UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

MÉMOIRE D'ÉTUDIANT

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

ÉTUDE APPROFONDIE D'UN MODÈLE D'ÉQUILIBRE POUR L'AFFECTATION DES PASSAGERS DANS LES RÉSEAUX DE TRANSPORT EN COMMUN

par

Michel GENDREAU

Thèse de Ph.D., département d'informatique et de recherche opérationnelle, dirigée par Michael Florian

Université de Montréal

Centre de recherche sur les transports - Publication #384

Mai 1984

Dépôt légal - 1er trimestre 1985

ERRATUM

Publication #384 : "Étude approfondie d'un modèle d'équilibre pour l'affectation des passagers dans les réseaux de transport en commun".

par : Michel Gendreau

Page 241, ligne 9, il faut lire $\eta_j = k_j \mu_j$,

plutôt que $\eta_j = \frac{k_j}{\mu_j}$.

Page 252, ligne 4, il faut lire $\frac{1}{\eta_j} = \frac{1}{k'_j \mu_j}$,

plutôt que $\eta_j = \frac{k'_j}{\mu_j}$.

Page 256, ligne 6, il faut lire $\frac{k_j}{\eta_j}$,

plutôt que $\frac{\eta_j}{k_j}$.

TABLE DES MATIÈRES

Table des illustrations viii

Sommaire xi

1. Introduction 1

2. Eléments communs aux modèles d'affectation des passagers . . . 9

2.1 Introduction 9

2.2 Nature spécifique du problème d'affectation des passagers 9

2.3 Modélisation statique 10

2.4 Modèles de flot 11

2.5 Représentation du réseau de transport 11

2.6 Demande de transport en commun et accès 14

2.7 Hypothèses relatives au coût des itinéraires 16

2.8 Attente aux arrêts 17

2.9 Différenciation des usagers. 17

3. Revue critique des principaux modèles d'affectation des passagers 19

3.1 Introduction 19

3.2 Les modèles simples d'affectation des passagers 19

3.2.1 Remarques générales 19

3.2.2 Algorithme de Dial (1967) 21

3.2.3 Approche de Fearnside et Draper (1971) 21

3.2.4 Méthode de Le Clercq (1972) 22

3.2.5 TRANSCOM (Chriqui, 1974 ; Chapleau, 1974) 22

3.2.6 Méthode d'Andréasson (1976) 23

3.2.7 NOPTS (Rapp et al., 1976) 24

3.2.8 Modèle de Spiess (1983) 25

3.3 Les modèles d'affectation multimodale 26

3.3.1 Remarques générales 26

3.3.2 TRANSEPT (Daly, 1973 ; Last et Leak, 1976) 27

3.3.3 Modèle de Florian (1977) 29

3.3.4 Modèle de Schéele (1976) 29

3.3.5 Modèle d'Abdulaal et LeBlanc (1979) 30

3.4 Les modèles de design de réseaux de transport en commun . 30

3.4.1 Remarques générales 30

3.4.2 Etude de Lampkin et Saalmans (1967) 31

3.4.3 Modèle de Holm (1973) 31

3.4.4 Modèle de Sharp (1973) 32

3.4.5 Modèle de Mandl (1980) 32

3.4.6 Approche de Hasselström (1981) 33

3.5 Commentaires sur les hypothèses de comportement 34

3.6 Itinéraires versus stratégies 37

3.7 Critiques relatives au traitement de la congestion 40

3.8 Remarques sur le calcul des temps d'attente 42

3.8.1 Temps d'attente aux arrêts simples 42

3.8.2 Temps d'attente aux arrêts multiples 44

3.8.3 Choix des distributions des intervalles inter-autobus 49

3.8.4 Le hasard et les arrivées des usagers aux arrêts . 50

3.9 Conclusion 52

4. Formulation du modèle d'affectation d'équilibre pour les réseaux sans arrêts multiples 54

4.1 Introduction 54

4.2 Particularités du modèle 55

4.2.1 Modèle d'équilibre 55

4.2.2 Déplacements à pied intégrés 55

4.2.3 Réseau de transport en commun 56

4.2.4 Coefficients de coûts 57

4.2.5 Coûts des déplacements à pied et des correspondances 58

4.2.6 Temps d'attente et capacité 58

4.2.7 Coût généralisé des sections de lignes 60

4.3 Description mathématique du réseau et notation 65

4.3.1 Remarques générales 65

4.3.2 Définitions préliminaires 66

4.3.3 Notation pour le réseau augmenté 67

4.3.4 Quelques précisions 69

4.3.5 Eclatement des arrêts 70

4.3.6 Notation pour le réseau éclaté 72

4.3.7 Variables de flot 74

4.3.8 Fonctions de coût sur les arcs 75

4.4 Affectations réalisables 76

4.5 Affectations d'équilibre 80

4.5.1 Interprétations diverses de l'optimisation égoïste. 80

4.5.2 Condition d'équilibre 81

4.5.3 Reformulation du principe de Wardrop. 84

4.5.4 Mesure de déséquilibre 84

4.5.5 Equilibre et flots totaux 89

4.5.6 Equilibre basé sur les "coûts d'aujourd'hui" . . . 90

4.6 Existence et unicité d'une solution d'équilibre 91

4.7 Importance des coûts non-constants pour les sections de lignes 91

5. Fonctions d'attente et quasi-files d'attente 96

5.1 Introduction 96

5.2 L'arrêt d'autobus en tant que quasi-file d'attente 97

5.3 Survol des travaux antérieurs sur les quasi-files d'attente avec service en groupes 99

5.4 Résultats exacts pour les quasi-files d'attente $Q\text{-}M/M^{[K]}/1$ 102

5.5 Bornes et approximations pour le temps d'attente moyen dans les quasi-files d'attente $Q\text{-}M/M^{[K]}/1$ en régime stationnaire 104

5.5.1 Bornes pour W_q dans les quasi-files d'attente $Q\text{-}M/M^{[K]}/1$ 104

5.5.2 Formules d'approximation pour W_q dans les quasi-files d'attente $Q\text{-}M/M^{[K]}/1$ 116

5.6 Extension des résultats de Pollaczek-Khinchine aux quasi-files d'attente Q-M/G/1 119

5.6.1 Introduction 119

5.6.2 Rappel des résultats de Pollaczek-Khinchine pour les files d'attente classiques M/G/1 119

5.6.3 Extension aux quasi-files d'attente Q-M/G/1 122

5.6.4 Interprétation intuitive 125

5.6.5 Lien avec les travaux de Finch 126

5.7 Temps d'attente moyen dans les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$ 126

5.7.1 Introduction 126

5.7.2 Rappel des résultats de Bailey et Downton 127

5.7.3 Lien avec les files d'attente $E_K/E_k/1$ 129

5.7.4 Approximation de base 135

5.7.5 Approximation hyperbolique sur la taille des groupes 161

5.7.6 Formule d'approximation simplifiée 166

5.8 Fonctions d'attente 171

5.9 Critique des hypothèses sous-jacentes au modèle 174

6. Formulation du modèle d'affectation d'équilibre pour les réseaux comportant des arrêts multiples 177

6.1 Introduction 177

6.2 Définitions préliminaires, notation additionnelle. 178

6.3 Caractérisation du comportement des usagers aux arrêts multiples et des stratégies 179

6.4 Définition du sous-modèle d'arrêt multiple 183

6.5 Eclatement des arrêts multiples 186

6.6 Variables de flot 193

6.7 Fonctions de coût sur les arcs 196

6.8 Affectations réalisables 197

6.9 Affectations d'équilibre 201

6.10 Quelques remarques 211

7. Le sous-modèle d'arrêt multiple 213

7.1 Introduction 213

7.2 Une classe de quasi-files d'attente pour les arrêts multiples 213

7.3 Séparation du problème de répartition et du problème d'attente 218

7.4 Sous-modèle de répartition 219

7.4.1 Introduction 219

7.4.2 Notation 219

7.4.3 Condition de stationnarité 220

7.4.4 Hypothèse fondamentale de répartition 222

7.4.5 Répartitions réalisables et répartitions d'équilibre 223

7.4.6 Existence et unicité de la répartition d'équilibre 228

7.4.7 Détermination des probabilités de répartition ; propriétés 237

7.5 Calcul des temps d'attente 240

7.5.1 Introduction 240

7.5.2 Distribution du minimum de n variables suivant des lois d'Erlang 241

7.5.3 Superposition de n processus d'Erlang 251

7.5.4 Pseudo-distribution induite pour les intervalles de service 260

7.5.5 Paramètres à utiliser dans le calcul des temps d'attente 263

7.6 Intégration du sous-modèle d'attente au modèle global d'affectation 266

7.7 Quelques remarques sur l'approche employée 270

8. Simulations d'arrêts 272

8.1 Introduction 272

8.2 Schéma général des simulations 273

8.3 Etude de l'impact de la variabilité de la taille des groupes 274

8.3.1 Aux arrêts simples 274

8.3.2 Aux arrêts multiples 277

8.4 Questions relatives au sous-modèle de répartition 281

8.5 Résultats relatifs au sous-modèle d'attente 285

8.6 Conclusion 286

9. Propriétés du modèle - Généralisations possibles 289

9.1 Introduction 289

9.2 Existence et unicité d'une solution d'équilibre pour le modèle global 289

9.3 Considérations algorithmiques 291

9.4 Généralisation du modèle aux classes multiples d'usagers 293

9.5 Modèle avec demandes variables 295

9.6 Critique générale du modèle 295

9.7 Quelques avenues de recherche pour le futur 297

Conclusion 299

Remerciements 300

Bibliographie 301

TABLE DES ILLUSTRATIONS

Figure 1. Représentations typiques de lignes à un et deux terminus 13

Figure 2. Réseau de l'exemple 3.1 38

Figure 3. Graphe typique de fonction $t_a(v_a)$ 60

Figure 4. Graphe typique de fonction $\overline{\Psi}_a(v_a)$ 61

Figure 5. Graphe typique de fonction $\widetilde{\Psi}_a(v_a)$ 62

Figure 6. Graphe typique de fonction Ψ 64

Figure 7. L'arrêt simple r en représentation non-éclatée 71

Figure 8. L'arrêt simple r en représentation éclatée 71

Figure 9. Réseau de l'exemple 4.1 92

Figure 10. Le sous-modèle d'arrêt multiple 186

Figure 11. Séparation des lignes 187

Figure 12. Ajout des arcs auxiliaires 188

Figure 13. Structure d'attente-répartition 189

Figure 14. L'arrêt multiple r dans le réseau augmenté 190

Figure 15. L'arrêt multiple r en représentation éclatée 191

Tableau 5.1. Bornes sur W_q pour une quasi-file d'attente $Q\text{-}M/M^{[K]}/1$ avec $\lambda=8$ et $\rho=0,8$ 116

Tableau 5.2. Valeurs de $\beta^-(K)$, $\beta^+(K)$, $\bar{\beta}(K)$ et $\tilde{\beta}(K)$ pour $\rho=0,8$ et $0,5$ et $K=2,5,10,50,100$ 118

Tableau 5.3 Comparaison des résultats obtenus par l'approximation de base avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$ avec taux d'arrivée unitaire 158

Tableau 5.4 Comparaison des résultats obtenus par l'approximation de base avec ceux déduits des tables de Hillier et Yu pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/1 avec taux d'arrivée unitaire 159

Tableau 5.5 Comparaison des résultats obtenus par l'approximation de base avec ceux de Roy pour le temps d'attente dans les quasi-file d'attente Q-M/$D^{[K]}$/1 avec taux d'arrivée unitaire 160

Tableau 5.6 Valeurs de $\tilde{Z}(k,\rho,K)$ et $Z^h(k,\rho,K)$ pour k=1,2,3,∞, K=1,2,3,6,12,24,100,∞ et ρ=0,8 et 0,9 163

Tableau 5.7 Comparaison des résultats obtenus par l'approximation hyperbolique avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/1 et Q-M/$D^{[K]}$/1 avec taux d'arrivée unitaire 165

Tableau 5.8 Valeurs de $\tilde{Z}(k,\rho,\infty)$ et $Z_\infty^S(k,\rho)$ pour k=1,2,3,4,50 et ρ variant de 0,0 à 1,0 par pas de 0,1 168

Tableau 5.9 Comparaison des résultats obtenus par l'approximation simplifiée avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/1 avec taux d'arrivée unitaire 170

Tableau 5.10 Correspondance entre les paramètres du modèle d'attente et les variables et paramètres du modèle d'équilibre pour un arrêt simple r donné . . . 172

Tableau 7.1 Correspondance entre les variables, paramètres et fonctions du modèle d'arrêt multiple et du modèle global d'affectation pour un arrêt multiple r donné 268

Tableau 8.1 Impact de la variabilité de la taille des groupes sur le temps d'attente simulé moyen pour des quasi-files d'attente Q-M/$E_k^{[K]}$/1 276

Tableau 8.2 Paramètres pour les séries 1 et 2 de simulations d'arrêt multiple 278

Tableau 8.3 Impact de la variabilité de la capacité sur les répartitions simulées pour un arrêt multiple r donné 279

Tableau 8.4 Impact de la variabilité de la capacité sur les temps d'attente simulés moyens pour un arrêt multiple donné 280

Tableau 8.5 Impact d'une variation uniforme de la taille des autobus sur les probabilités de répartition 282

Tableau 8.6 Comparaison des résultats obtenus par simulation avec ceux produits par les modèles de répartition au prorata des fréquences et des capacités résiduelles pour deux exemples d'arrêt multiple . . . 284

Tableau 8.7 Comparaison des résultats obtenus par simulation avec ceux produits par le sous-modèle d'attente pour un arrêt multiple 287

SOMMAIRE

Dans cette thèse, nous nous intéressons au problème de l'affectation des passagers dans les réseaux de transport en commun. Ce problème consiste à déterminer, à partir d'un ensemble de demandes de transport (supposées connues et fixes) et des paramètres de service du système de transport en commun (supposés eux aussi connus et fixes) les stratégies utilisées par les usagers pour se rendre de leur origine à leur destination et la charge qui en résultera sur les différentes portions des lignes de transport.

L'étude des modèles antérieurs pour résoudre ce problème ayant montré qu'il n'existait pas vraiment de modèle qui tienne compte de manière adéquate des effets de la congestion du système de transport en commun sur l'affectation des passagers, nous proposons un nouveau modèle dont l'objectif premier est de combler ce vide.

Le modèle proposé est un modèle d'équilibre statique fondé sur l'hypothèse d'optimisation égoïste ("user-optimization") pour le comportement des usagers. Les effets de la congestion sur la répartition des usagers sont représentés dans le modèle à deux niveaux : en pénalisant par des coûts perçus de parcours plus élevés les déplacements effectués sur les lignes fortement chargées, et, surtout, en utilisant pour le calcul des temps d'attente aux arrêts des fonctions qui dépendent du nombre de personnes voulant prendre l'autobus et de la disponibilité de places libres.

Les fonctions d'attente aux arrêts simples (arrêts desservis par une seule ligne) sont obtenues à partir d'une famille d'approximations pour le temps d'attente moyen en régime stationnaire dans les quasi-files d'attente ("dispatch queues") avec service en groupes et temps de service distribués selon des lois d'Erlang. Nous avons effectué un certain nombre de comparaisons entre des valeurs exactes connues du temps d'attente et les valeurs obtenues à partir des formules d'approximation ; ces comparaisons montrent que les erreurs d'approximation sont tout à fait minimes, bien que le calcul des valeurs approchées soit beaucoup plus facile que celui des valeurs exactes.

L'épineuse question des arrêts multiples (arrêts desservis par plusieurs lignes) est abordée par le biais d'une légère reformulation de la notion de stratégie proposée par Spiess (1983) : nous supposons dans le modèle que chaque usager se présentant à un arrêt multiple choisit a priori un sous-ensemble de lignes attrayantes et qu'il montera à bord du premier autobus d'une de ces lignes dans lequel il trouvera une place libre. Après avoir montré que cette hypothèse permet de définir un état d'équilibre global pour le système de transport en commun, nous définissons à partir de celle-ci un sous-modèle d'arrêt multiple qui nous permet de calculer de manière approchée la répartition des usagers entre les lignes attrayantes pour eux et leur temps d'attente.

Pour évaluer le sous-modèle d'arrêt multiple, nous avons effectué une série de simulations qui indiquent que celui-ci produit des résultats satisfaisants pour les arrêts assez achalandés.

La fin de la thèse est consacrée à des questions relatives aux propriétés du modèle global (existence et unicité d'une affectation d'équilibre), à des considérations algorithmiques, à des généralisations possibles du modèle proposé (classes multiples d'usagers, demandes variables), ainsi qu'à une critique générale de celui-ci.

CHAPITRE PREMIER

INTRODUCTION

Au cours des dernières années, l'utilisation de la recherche opérationnelle est devenue pratique courante dans la résolution de problèmes de planification complexes dans de nombreux domaines. Un des champs d'application privilégiés des modèles et des techniques de la recherche opérationnelle est celui des transports, particulièrement celui du transport en commun. Il est, en effet, souvent essentiel pour la planification à court et à moyen termes du développement du système de transport en commun de pouvoir déterminer comment celui-ci se comportera dans certaines conditions. Les modèles d'affectation des passagers permettent de satisfaire en partie ce besoin. Plus précisément, ces modèles essaient de répondre le plus exactement possible à la question suivante :

Etant donnés une configuration du réseau de transport en commun (localisation des lignes, arrêts, stations), un niveau de service (capacité des véhicules, fréquences des autobus ou des rames de métro sur lignes), une politique d'opération (tarification, règles de correspondance) et un niveau de demande de transport en commun (nombre de déplacements des origines vers les destinations), quelle sera l'utilisation que les usagers feront du réseau et quelle charge en résultera sur les différentes portions de celui-ci?

Idéalement, il faudrait pour répondre à une telle question un

modèle dynamique qui tiendrait compte des aspects temporels du déplacement des usagers dans le réseau de transport. Malheureusement, les modèles dynamiques développés jusqu'à maintenant ne nous permettent pas de traiter de manière suffisamment efficace les réseaux de transport en commun réels. Il faut donc s'en tenir à des modèles statiques par lesquels on tente de déterminer l'état moyen du système étudié pendant la période donnée.

Dans une première analyse, on peut ramener l'utilisation par les usagers du réseau de transport en commun à une question de choix d'itinéraires. Il s'agit donc de déterminer quels itinéraires seront choisis par les usagers pour se déplacer des origines vers les destinations.

Plusieurs facteurs influent sur la perception qu'ont les usagers des différents itinéraires qui s'offrent à eux : temps total de parcours, temps d'attente, distance à parcourir entre l'origine ou la destination et des points d'accès au réseau, encombrement des lignes utilisées, etc. Dans le cadre d'un modèle d'affectation, il est nécessaire de s'en tenir en pratique à un seul critère de choix. Pour définir ce critère, le modélisateur doit utiliser l'une ou l'autre des approches suivantes :

a) déterminer le facteur de choix prépondérant pour les usagers et en faire le critère de choix pour le modèle;
b) construire un critère de coût généralisé basé sur un certain nombre des facteurs de choix, pondérés de manière à refléter leur importance relative pour l'usager moyen du système.

Dans un cas comme dans l'autre, on obtient une mesure permettant de comparer les itinéraires; par la suite, nous l'appellerons tout simplement le coût des itinéraires.

Une fois posé le critère de coût, il faut caractériser le comportement macroscopique des usagers en fonction de celui-ci. Ceci est fait en posant une hypothèse de comportement des usagers. En 1952, Wardrop a énoncé, sous une forme légèrement différente, les deux principes suivants pour le transport routier :

1) Pour chaque paire origine-destination, le coût des itinéraires utilisés est égal et il est moindre que celui des itinéraires non-utilisés.
2) Le coût moyen de transport est minimal.

Le premier principe correspond à la situation dans laquelle chaque automobiliste essaie de minimiser son coût de transport personnel sans se préoccuper des autres usagers; ce comportement est appelé "optimisation égoïste" (en anglais, "user-optimization").

Le second principe, quant à lui, correspond à la situation dans laquelle les automobilistes coopèrent dans le choix de leur itinéraire (sous la direction d'une autorité centrale, par exemple) de façon à minimiser le coût total de transport de l'ensemble des usagers; c'est ce qu'on appelle l'"optimisation coopérative" (en anglais, "system-optimization").

Toujours dans le contexte du transport routier, Dial (1971) a proposé une hypothèse différente basée sur le modèle logistique multinomial (originellement développé en théorie du choix aléatoire). Pour chaque paire origine-destination, après avoir déterminé un ensemble d'itinéraires

acceptables (en termes de coût généralisé), on répartit les usagers entre ceux-ci de manière "probabiliste" selon les spécifications du modèle logistique. Nous commenterons cette hypothèse dans le prochain chapitre.

De manière générale, on peut reprendre les hypothèses de comportement utilisées en transport routier pour le transport en commun. Il faut cependant les modifier légèrement dans le cas où le réseau étudié comporte des arrêts multiples, c'est-à-dire des arrêts auxquels passent plusieurs lignes d'autobus distinctes. En effet, un usager qui se présente à un arrêt multiple peut choisir d'y attendre un autobus d'une ligne donnée, mais d'autres choix s'offrent aussi à lui. Il peut, entre autres, élargir son choix à un sous-ensemble de lignes et prendre la première place disponible sur l'une de celles-ci. Un comportement de ce genre aura tendance à diminuer de beaucoup les temps d'attente et donc à être plus efficace du point de vue des usagers. Il est essentiel de remarquer que dans ce cas, on ne peut plus vraiment parler de l'itinéraire d'un usager de son origine vers sa destination. En effet, à chaque arrêt multiple où le choix d'un usager ne se restreint pas à une seule ligne, le déplacement de celui-ci dans le réseau peut se poursuivre de plusieurs façons. Pour résoudre cette difficulté, on définit la notion de stratégie des usagers (voir Spiess, 1983). De manière large, on dira qu'une stratégie est un ensemble de règles permettant à un usager de se rendre de son origine à sa destination, en fonction d'un certain nombre d'événements aléatoires pouvant se produire au cours de son déplacement.

A cause des éléments aléatoires dans le déplacement des usagers, il n'existe pas, à proprement parler, de coût pour une stratégie. C'est

plutôt en termes d'espérance (mathématique) du coût de transport qu'il faut reformuler les hypothèses de comportement des usagers. On obtient alors :

a) Pour l'optimisation égoïste : chaque usager choisit sa stratégie de manière à minimiser l'espérance de son coût de transport propre.

b) Pour l'optimisation coopérative : les usagers choisissent leur stratégie de façon à minimiser l'espérance totale du coût de transport de l'ensemble des usagers.

Une autre question fondamentale dans un modèle d'affectation des passagers est celle du traitement de la congestion du système de transport en commun. Il est clair qu'aux heures de pointe, la congestion peut devenir assez importante sur certaines lignes. Dans ce cas, les usagers ne peuvent pas toujours monter à bord du premier autobus qui se présente à leur arrêt, ce qui fait augmenter sensiblement le temps d'attente. Ceci a comme conséquence de rendre moins attrayantes les lignes les plus surchargées et de pousser les usagers vers d'autres lignes. La congestion peut donc avoir un effet marqué sur la répartition des usagers entre les lignes, et il faut en tenir compte dans un modèle d'affectation des passagers si l'on veut pouvoir traiter adéquatement des réseaux congestionnés.

Comme nous le verrons au chapitre suivant, le phénomène de congestion du système de transport en commun est, à toutes fins pratiques, ignoré dans la quasi-totalité des modèles existants d'affectation des passagers. Il nous paraît essentiel de combler ce vide. Dans notre mémoire de maîtrise (Gendreau, 1977), nous avions proposé un modèle

d'affectation qui tenait compte des effets de la congestion, mais qui ne pouvait traiter les réseaux comportant des arrêts multiples. Dans cette thèse, nous reprenons ce modèle en y ajoutant les éléments nécessaires au traitement des arrêts multiples.

Dans ce nouveau modèle, le critère de coût retenu est une somme pondérée du temps d'accès au réseau, du temps d'attente aux arrêts et du temps passé à bord d'autobus, avec un facteur de correction pour pénaliser les déplacements sur des lignes fortement chargées. Ce facteur représente l'inconfort pour les usagers de voyager dans des autobus presque pleins. L'hypothèse de comportement utilisée est celle de l'optimisation égoïste (reformulée par rapport aux stratégies). Les effets de la congestion sur la répartition des usagers sont représentés dans le modèle par les deux moyens suivants :

a) les coûts supplémentaires qui sont appliqués aux déplacements sur les lignes fortement chargées (facteur ayant une influence marginale sur l'affectation);

b) les fonctions de temps d'attente aux arrêts qui dépendent du nombre de personnes voulant prendre l'autobus à un arrêt donné et de la disponibilité de places libres sur les lignes desservant cet arrêt (facteur ayant un impact majeur sur la répartition).

Les principaux éléments originaux du modèle proposé sont, à notre avis :

a) le traitement des effets de la congestion par le biais des temps d'attente aux arrêts;

b) le fait que les fonctions d'attente aux arrêts utilisées sont des approximations basées sur des résultats tirés de la théorie des quasi-files d'attente (en anglais, "dispatch queues");

c) l'utilisation d'un sous-modèle "microscopique" très détaillé pour les arrêts multiples;

d) la condition d'équilibre du modèle qui représente une extension nouvelle du premier principe de Wardrop.

Nous présentons aussi aux chapitres quatrième et sixième de nouveaux résultats relatifs aux quasi-files d'attente et aux superpositions de processus d'Erlang.

La thèse est séparée en quatre grandes parties : après une brève description des éléments communs à tous les modèles d'affectation des passagers (chapitre deuxième) et une revue critique des principaux modèles existants (chapitre troisième), nous présentons le modèle d'affectation en présence de congestion dans sa version la plus simple, c'est-à-dire pour les réseaux sans arrêts multiples (chapitres quatrième et cinquième); la troisième partie (chapitres sixième, septième et huitième) porte sur les modifications qu'il faut apporter au modèle pour le traitement des arrêts multiples; finalement, nous étudions au chapitre neuvième quelques-unes des propriétés du modèle dans ses deux versions et nous indiquons quelques avenues de recherche pour l'avenir. Dans chacune des deux parties centrales, le premier chapitre (chapitres quatrième et sixième respectivement) est consacré à la description globale du modèle en utilisant la théorie des flots dans les réseaux, et le second (chapitres (cinquième et

septième respectivement) aux sous-modèles "microsocpiques" d'arrêts et à la dérivation de fonctions d'attente. Quant au chapitre huitième, nous y indiquons quelques résultats de simulations qui valident en partie le sous-modèle d'arrêt multiple et les approximations utilisées pour les fonctions d'attente.

CHAPITRE DEUXIÈME

ÉLÉMENTS COMMUNS AUX MODÈLES D'AFFECTATION DES PASSAGERS

2.1 Introduction

Tous les modèles d'affectation des passagers, que ce soient ceux qui ont été proposés par divers auteurs depuis une quinzaine d'années, ou ceux que nous développons dans cette thèse, possèdent un certain nombre de traits communs qui découlent de la nature même du problème à résoudre. Dans le but d'alléger la présentation des chapitres qui suivent, nous avons regroupé ici tous ces éléments communs.

2.2 Nature spécifique du problème d'affectation des passagers

Dans le problème d'affectation du trafic routier, il y a une équivalence directe entre les déplacements des usagers (qui sont en fait dans ce cas des automobilistes) et les déplacements de véhicules. Ceci a pour conséquence de ramener le problème d'affectation du trafic à une question de détermination d'itinéraires de véhicules qui est relativement simple à résoudre. Dans le problème d'affectation des passagers d'un réseau de transport en commun, cette équivalence ne tient plus : les déplacements des usagers, même s'ils sont contingents aux déplacements des autobus ou des rames de métro sur les lignes du réseau, ne correspondent jamais totalement à ceux-ci. Il faut en effet que les usagers fassent accès au réseau, c'est-à-dire qu'ils se rendent de leur origine à un arrêt d'autobus ou une station de métro, qu'ils attendent un véhicule, qu'ils effectuent

éventuellement des correspondances (ce qui implique à chaque fois une nouvelle attente) et finalement qu'ils quittent le réseau de transport en commun pour se diriger à pied vers leur destination. Tout ceci complique évidemment énormément le problème et fait qu'il ne peut être traité comme une simple variante du problème d'affectation du trafic pour lequel il existe d'excellents modèles.

2.3 Modélisation statique

Dans sa nature première, le problème d'affectation des passagers est essentiellement dynamique : les usagers quittent leur origine à un moment précis dans le temps, et cela influe sur le reste de leurs déplacements. Pourtant, tous les modèles développés jusqu'à ce jour négligent cet aspect et tentent plutôt de déterminer l'état moyen du système pendans la période d'intérêt, c'est-à dire la charge moyenne d'usagers sur les différentes portions du réseau et les coûts moyens de parcours par itinéraire (ou par stratégie) entre les origines et les destinations. Il existe deux raisons expliquant cet état de fait :

a) Les modèles d'affectation des passagers étant utilisés pour l'étude de réseaux de transport de taille assez considérable, l'état moyen du système pendant la période étudiée (heure de pointe) représente probablement l'ensemble d'informations le plus pertinent pour servir de base aux décisions relatives au développement futur du réseau considéré; avec un modèle plus détaillé, on risquerait sans doute d'ensevelir le planificateur sous une masse de détails inutiles qui le gêneraient dans son travail, au lieu de lui faciliter la tâche.

b) Un modèle mathématique plus détaillé serait d'une taille et d'une complexité telles qu'il nous serait impossible de l'utiliser pour traiter des réseaux de transport réels.

2.4 Modèles de flot

Les modèles d'affectation sont tous explicitement ou implicitement des modèles de flot (voir Ford et Fulkerson, 1962) dans un réseau (au sens mathématique du terme) construit à partir du réseau de transport en commun. Dans ce réseau, chaque arc représente soit un déplacement à pied ou en autobus, soit l'attente à un arrêt, et le coût (ou longueur) qui lui est associé est égal au coût généralisé de ce déplacement ou de cette attente. Les déplacements d'usagers suivant un itinéraire (ou chemin) sont assimilés à l'établissement d'un flot statique entre l'origine de cet itinéraire et sa destination. La valeur du flot total sur chaque arc correspond au nombre moyen d'usagers qui l'empruntent par unité de temps.

Il est important de souligner ici que, dans chaque modèle, le réseau sur lequel sont définis les flots d'usagers possède un certain nombre de particularités qui découlent soit de l'hypothèse de comportement choisie, soit d'autres aspects de la modélisation; il nous sera néanmoins possible de dégager dans les sections suivantes un certain nombre de points communs entre ces réseaux.

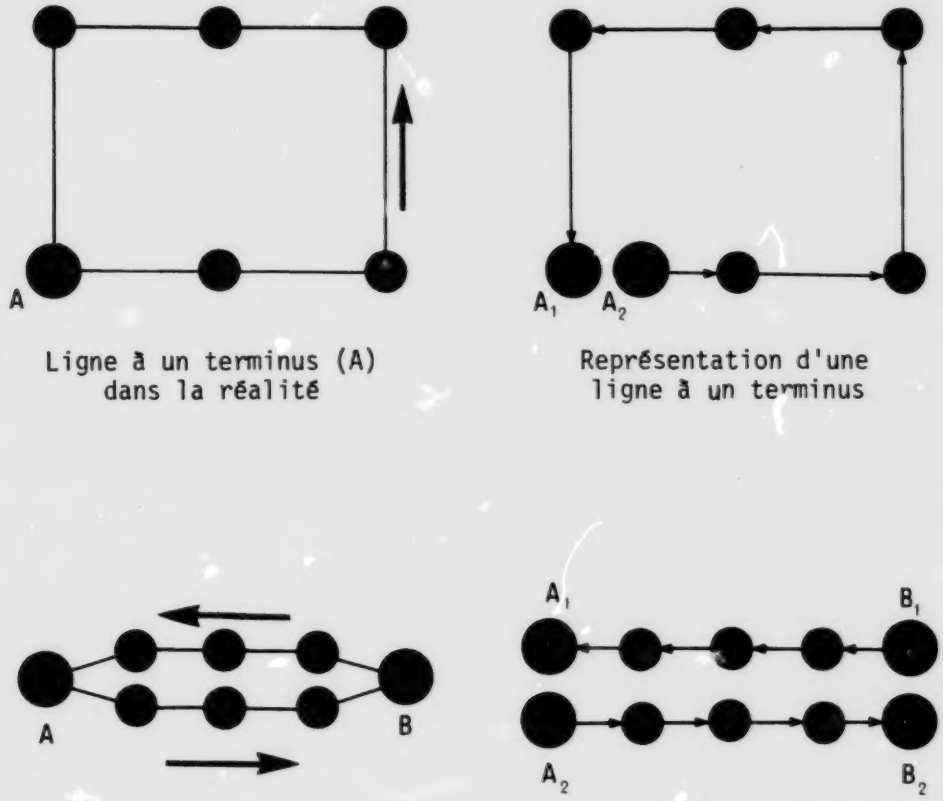
2.5 Représentation du réseau de transport

De manière générale, on décrit le réseau de transport en commun sous la forme d'un graphe orienté (appelé graphe de base du réseau) dans

lequel les sommets correspondent aux arrêts (ou stations) et les arcs aux sections de lignes de transport. Malheureusement, dans la définition de ce graphe, on ne peut se contenter d'utiliser directement une carte du réseau indiquant la localisation des lignes et des arrêts en établissant les correspondances arrêts-sommets et sections de lignes-arcs. En effet, dans les grandes villes, les réseaux comportent généralement tellement de points d'accès qu'il est à peu près impossible de les inclure tous dans un modèle d'affectation des passagers. Pour ramener le nombre d'arrêts utilisés dans le modèle à un chiffre "raisonnable", on choisit donc soit de n'en conserver qu'un sous-ensemble (ceux qui paraissent les plus significatifs), soit de procéder à une agrégation par laquelle on remplace plusieurs points d'accès réels par un arrêt fictif. Dans la suite, quand nous parlerons de l'ensemble des arrêts du réseau, nous ferons toujours référence, sauf mention explicite à l'effet contraire, à l'ensemble des arrêts choisis ou des arrêts agrégés.

Une fois déterminé l'ensemble des arrêts du réseau, il faut préciser la représentation des lignes d'autobus. Ordinairement, celles-ci forment des circuits fermés avec un ou deux terminus, points auxquels tous les usagers à bord de l'autobus doivent descendre. Pour modéliser adéquatement le rôle des terminus, on est généralement forcé de les dédoubler dans la représentation des lignes : un des sommets, appelé terminus d'arrivée, est le point auquel les usagers descendent, et l'autre, appelé terminus de départ, celui auquel d'autres passagers peuvent monter dans l'autobus maintenant vide. Pour les lignes qui relient deux terminus, il faut utiliser deux lignes dans les modèles : une pour représenter la portion "aller"

du trajet et l'autre pour la portion "retour". La figure 1 nous montre des représentations typiques de lignes à un et deux terminus.



Ligne à un terminus (A) dans la réalité

Représentation d'une ligne à un terminus

Ligne à deux terminus (A,B) dans la réalité

Représentation d'une ligne à deux terminus

Figure 1 : Représentations typiques de lignes à un et deux terminus.

Dans la figure 1, les sommets A_1 et B_2 sont des terminus d'arrivée et les sommets A_2 et B_1 des terminus de départ. On remarque que les lignes ne forment pas des circuits dans la représentation, mais plutôt des chemins débutant à un terminus de départ et se terminant à un terminus d'arrivée.

2.6 Demandes de transport en commun et accès

Il est impossible, même pour une ville de petite taille, de localiser précisément l'origine et la destination de chaque usager potentiel du système de transport en commun. Il faut donc travailler avec un certain niveau d'agrégation.

Au départ, la région qui fait l'objet de l'étude est divisée en un certain nombre de zones; ce découpage est fait de façon à ce que les zones soient aussi homogènes que possible du point de vue de l'utilisation des sols. On estime ensuite, pour chaque paire de zones, la demande totale de transport de l'une vers l'autre soit à partir d'une enquête des ménages, soit à partir d'un modèle de génération des déplacements ("trip generation model"). Cela fait, on effectue une répartition modale pour chaque paire (de zones) origine-destination, c'est-à-dire que, en se basant sur des estimés des coûts de transport de l'origine vers la destination par les divers modes disponibles (automobile, transport en commun, marche, etc.) et sur certaines autres données, on répartit la demande totale entre les modes de transport. On obtient donc en définitive pour chaque paire O-D, une demande de transport en commun. Cette demande peut être exprimée sous deux formes : soit par un nombre unique (la demande ne varie pas en fonction des résultats de l'affectation des passagers et l'on parle alors de demande fixe), soit par une fonction qui donne le niveau de demande de transport en commun relativement au coût généralisé de transport de l'origine vers la destination (on parle alors de demande variable). Dans le cas des demandes fixes, l'ensemble de ces demandes forme ce qu'on appelle la matrice origine-destination (ou matrice O-D).

Les modèles d'affectation des passagers résolvent généralement le problème avec demandes fixes, nous montrerons cependant au chapitre neuvième que le modèle que nous proposons peut être modifié pour résoudre le problème avec demandes variables.

Pour être en mesure d'utiliser les demandes de transport, il faut pouvoir les situer géographiquement par rapport au réseau de transport en commun. Cela est fait en définissant pour chaque zone un point que l'on appelle le centroïde de cette zone; c'est en ce point que l'on suppose que se concentrent toutes les demandes de transport relatives à la zone.

Pour représenter l'accès au réseau de transport, on relie, dans la plupart des modèles, chaque centroïde à certains des arrêts situés à proximité de celui-ci par des arcs de marche fictifs. Les arrêts auxquels on relie un centroïde donné sont choisis en fonction de la configuration particulière du réseau de transport aux alentours de ce centroïde et des caractéristiques socio-économiques de la zone correspondante. Certains modèles traitent le problème de l'accès de manière légèrement différente, mais cela n'a aucune importance pour les fins de la revue du chapitre suivant.

Le graphe obtenu par l'ajout au graphe de base du réseau des centroïdes, des arcs d'accès (arcs reliant centroïdes et arrêts), d'arcs de correspondance (arcs reliant des arrêts assez proches pour permettre des correspondances entre les lignes) et éventuellement d'arcs de marche reliant les centroïdes (si l'on permet aux usagers d'effectuer leurs déplacements à pied) sera appelé graphe augmenté du réseau; pour tout modèle

d'affectation, c'est une donnée fondamentale, bien que cela ne correspond pas tout à fait au réseau définitif sur lequel sont définis les flots (dans chaque modèle, on ajoute des arcs ou des sommets jouant un rôle spécifique à celui-ci).

2.7 Hypothèses relatives au coût des itinéraires

Dans les modèles, on fait deux hypothèses importantes relativement au coût des itinéraires : ce sont celle de l'homogénéité des usagers par rapport aux coûts et celle de l'additivité des coûts.

L'hypothèse d'homogénéité des usagers par rapport aux coûts consiste à supposer que tous les usagers accordent la même importance relative aux différentes composantes de leur déplacement (par exemple, pour tous les usagers, le temps d'attente aux arrêts est deux fois plus déplaisant que le temps passé à bord de l'autobus). C'est une hypothèse simplificatrice qui n'est peut-être pas toujours justifiable en pratique. Dans le cas où elle ne l'est pas, il faut séparer les usagers en un certain nombre de classes distinctes correspondant, elles, à des groupes d'usagers homogènes et modifier en conséquence le modèle global, comme nous le verrons au chapitre neuvième.

L'hypothèse d'additivité des coûts consiste, quant à elle, à supposer que le coût total d'un itinéraire est égal à la somme des coûts des éléments qui le composent (déplacements à pied et en autobus, attente). Quant on inclut le prix de passage exigé dans le coût des itinéraires, l'hypothèse n'est pas vérifiée si on a une tarification non-additive. Il faut alors apporter des modifications "ad hoc" au modèle pour tenir compte de ce fait.

2.8 Attente aux arrêts

L'attente aux arrêts joue un grand rôle dans l'affectation des passagers, car c'est souvent la composante du déplacement dans un réseau de transport en commun que les usagers détestent le plus. Les modèles diffèrent sensiblement dans leur traitement exact de la question, mais, sauf dans de très rares exceptions, le temps d'attente aux arrêts simples (c'est-à-dire aux arrêts auxquels passe une seule ligne) est toujours relié à l'intervalle moyen entre deux autobus successifs (en anglais, "headway") sur la ligne qui les dessert. Aux arrêts multiples, le calcul du temps d'attente requiert une procédure beaucoup plus complexe qui varie avec chaque modèle.

L'intégration des coûts d'attente dans le réseau sur lequel sont définis les flots d'usagers implique le dédoublement implicite ou explicite des sommets qui représentent les arrêts et l'ajout (toujours implicite ou explicite) d'un arc représentant l'attente pour chaque arrêt (le coût associé à cet arc étant le coût d'attente à l'arrêt).

Notons aussi que le temps d'attente à un arrêt peut être, selon les modèles, dépendant ou indépendant du nombre d'usagers montant à cet arrêt et du nombre de places libres dans les autobus qui y passent.

2.9 Différenciation des usagers

Nous avons vu que dans la formulation du problème d'affectation des passagers comme problème de flot, le flot sur un arc du réseau représente le nombre moyen de passagers qui empruntent cet arc par unité de temps. En fait, on ne peut se contenter d'une seule variable de flot par

arc, car les usagers ne sont pas "interchangeables" : ils se dirigent tous d'une origine donnée vers une destination donnée; il faut donc les différencier d'une manière ou d'une autre.

La solution la plus naturelle à ce problème consiste à définir pour chaque arc une variable de flot par paire origine-destination, ce qui permet de bien distinguer les usagers; elle a comme désavantage de faire grossir énormément le nombre total de variables dans le modèle. Heureusement, il existe des manières plus économiques de différencier adéquatement les usagers. On peut en effet remarquer qu'une fois qu'un usager a quitté son origine, la seule chose qui importe, c'est sa destination. Ceci va nous permettre de ne différencier les usagers que par leur destination si on le désire. Dans les modèles d'affectation du trafic routier, on procède généralement à un renversement implicite du même raisonnement en distinguant les automobilistes par origine; c'est aussi la solution qui est utilisée dans la majorité des modèles d'affectation des passagers.

En résumé, on peut donc différencier les passagers en définissant pour chaque arc du réseau une variable de flot soit par paire O-D, soit par destination, soit par origine. Dans tous les cas, on aura plusieurs variables de flot pour chaque arc et l'on dira que le modèle final obtenu en est un de flot de plusieurs "produits" ("multi-commodity flow model"), un "produit" correspondant soit à une paire O-D, soit à une destination, soit à une origine.

CHAPITRE TROISIÈME

REVUE CRITIQUE DES PRINCIPAUX MODÈLES D'AFFECTATION DES PASSAGERS

3.1 Introduction

Dans ce chapitre, nous présentons rapidement quelques modèles d'affectation des passagers qui nous paraissent particulièrement représentatifs de ce qui s'est fait dans ce domaine depuis une quinzaine d'années. A part les modèles simples d'affectation des passagers (c'est-à-dire des modèles qui ne s'adressent qu'à cette question), nous considérons deux classes de modèles qui résolvent des problèmes plus vastes : les modèles d'affectation multimodale et les modèles de "design" de réseaux de transport en commun; dans chacun de ces deux cas, l'affectation des passagers constitue un sous-problème important du problème traité.

Après une première partie de chapitre essentiellement descriptive, nous ferons une critique globale des modèles présentés; cette critique s'articulera autour des quatre thèmes suivants :

a) hypothèses de comportement utilisées,

b) itinéraires versus stratégies,

c) traitement de la congestion,

d) calcul des temps d'attente aux arrêts.

3.2 Les modèles simples d'affectation des passagers

3.2.1 Remarques générales

Les sept modèles qui suivent ont été élaborés entre 1967 et 1983

pour tenter de résoudre le problème d'affectation des passagers aussi simplement et aussi efficacement que possible.

Dans tous les cas, le problème de congestion du système de transport en commun a été ignoré par les auteurs (c'est-à-dire qu'ils ont supposé que la capacité du réseau était amplement suffisante pour rendre la congestion négligeable). Il faut noter que quand on néglige le phénomène de congestion, il n'y a pas de différence entre l'optimisation égoïste et l'optimisation coopérative. Cela vient du fait qu'il n'y a pas d'interaction entre les usagers : si chacun choisit son itinéraire (ou sa stratégie) optimal, on obtient une affectation qui est optimale d'un point de vue global.

A l'exception de Spiess, les auteurs ne se sont pas attaqués au problème des arrêts multiples, mais ont plutôt essayé de résoudre celui des "lignes communes". Par "lignes communes", on entend la superposition sur une portion de leur parcours de deux ou plusieurs lignes d'autobus. En fait, les situations de lignes communes correspondent à un sous-ensemble des situations d'arrêt multiple, car si les lignes communes engendrent des arrêts multiples, il y a des arrêts multiples à partir desquels les lignes impliquées divergent. Qui plus est, le auteurs se sont limités dans leur traitement des lignes communes à ne considérer que les cas où l'arrêt auquel vont descendre les usagers appartient à la même portion de lignes communes que celui auquel ils montent dans l'autobus. Cela a pour effet de ramener, à toutes fins pratiques, le choix de stratégie des usagers à un choix d'itinéraire. Nous reviendrons plus en détail sur cette question à la section 3.6.

En terminant, il convient de souligner que la majorité des modèles présentés dans cette section ne méritent pas vraiment ce nom, leurs auteurs se bornant le plus souvent à ne définir qu'un algorithme permettant de calculer des affectations de passagers, sans expliciter les hypothèses de modélisation sous-jacentes.

3.2.2 Algorithme de Dial (1967)

Dès 1967, Dial a présenté une méthode de calcul d'itinéraires à coût minimum pour les réseaux de transport en commun. Cette méthode est basée sur une modification de l'algorithme de plus court chemin de Moore (1957). Le temps d'attente à un arrêt est égal à la moitié de l'intervalle interautobus moyen pour la ligne qui le dessert (demie de l'inverse de la fréquence de cette ligne). Les sections de lignes communes sont remplacées par des "super-arcs" ("trunk-line links" ou TLL). Sur les TLL, le temps de parcours est constant; c'est donc dire que si des lignes ayant des vitesses différentes ont une portion commune, il faut négliger cet aspect dans le modèle. D'autre part, pour les arrêts de TLL, le temps d'attente est égal à la moitié de l'inverse de la fréquence combinée des lignes du TLL (cela ne correspond pas à la moitié de l'intervalle interautobus moyen à cet arrêt comme nous le verrons à la section 3.8).

3.2.3 Approche de Fearnside et Draper (1971)

Fearnside et Draper ont suggéré d'utiliser les logiciels d'affectation du trafic routier (sans congestion) pour l'affectation des passagers. Ces logiciels comportant un mécanisme particulier pour pénaliser les virages aux intersections, il devient relativement facile de tenir compte du

temps d'attente aux arrêts par le biais de celui-ci (si le réseau est codé de manière astucieuse). Ce temps d'attente est, comme dans le modèle de Dial, fixé à la moitié de l'intervalle interautobus moyen. A cause de la transformation du problème d'affectation des passagers en problème d'affectation du trafic, il est impossible, dans cette approche, de tenir compte des lignes communes qui doivent être représentées comme des lignes totalement distinctes.

3.2.4 Méthode de Le Clercq (1972)

La méthode d'affectation des passagers de Le Clercq consiste aussi en une recherche des itinéraires à coût minimum; cependant, contrairement à Dial et à Fearnside et Draper qui utilisent une représentation arcs-sommets pour le réseau de transport en commun, Le Clercq a recours à une représentation lignes-sommets qui lui permet de résoudre le problème des lignes communes avec élégance. Pour ce qui est des temps d'attente, son approche est identique à celle de Dial. Les lignes communes sont intégrées directement au modèle par le biais de la représentation lignes-sommets,mais entre deux sommets d'une section commune, le temps de parcours est le même quelle que soit la ligne utilisée. Les usagers qui utilisent des sections de lignes communes sont répartis entre ces lignes proportionnellement aux fréquences respectives. Bien qu'à première vue, cette méthode semble assez différente de celle de Dial, les résultats qu'elles produisent sont équivalents.

3.2.5 TRANSCOM (Chrigui, 1974; Chapleau, 1974)

TRANSCOM est un logiciel d'affectation des passagers conçu pour

traiter de manière simple et efficace les réseaux de transport en commun de grande taille. Deux algorithmes sont disponibles pour le cheminement des usagers : le premier ne considère que la ligne la plus rapide entre deux points, tandis que le second tient compte de l'existence de lignes communes. Le traitement des lignes communes est, à notre avis, l'aspect le plus intéressant de TRANSCOM. Pour la première fois, en effet, on considère le cas où sur une section commune, les lignes peuvent avoir des vitesses différentes. Dans ce cas, il n'est pas toujours optimal de prendre le premier autobus qui se présente à l'arrêt. Chriqui et Robillard (1975) ont défini de manière générale le problème du choix du sous-ensemble optimal de lignes pour les portions communes et ils ont développé un algotithme qui résout ce problème dans le cas où les intervalles interautobus sont distribués selon des lois exponentielles. Leur algorithme permet aussi d'obtenir une bonne solution au problème si les intervalles interautobus suivent des distributions différentes. Dans TRANSCOM, on suppose que les distributions des intervalles interautobus sont exponentielles, ce qui permet non seulement de déterminer le sous-ensemble optimal de lignes, mais aussi de calculer exactement les temps d'attente moyens aux arrêts et la répartition des usagers entre les lignes communes.

3.2.6 Méthode d'Andréasson (1976)

Il est assez difficile de voir tout à fait précisément le problème qu'essayait de résoudre Andréasson. De manière globale, on peut dire qu'il s'agit d'une question d'affectation des passagers et que le critère de choix prépondérant est la minimisation du nombre de correspondances. Dans le cas où il existe plusieurs itinéraires ayant le même nombre de

correspondances, un ensemble de règles heuristiques plus ou moins complexes selon que l'on a une exploitation à horaires ou sans horaires, permet de répartir les usagers entre eux-ci. Il est certain que le manque de précision dans l'exposition de la méthode proposée nous empêche de nous faire une idée très claire sur sa validité.

3.2.7 NOPTS (Rapp et al., 1976)

NOPTS est un logiciel d'affectation des passagers basé sur l'hypothèse de comportement de Dial. Le temps d'attente aux arrêts est égal à la moitié de l'intervalle interautobus moyen. Pour les lignes communes, on utilise une approche qui dans son esprit est presque identique à celle de Le Clercq, mais dont l'application peut produire des résultats différents. En effet, à cause de certaines particularités de l'hypothèse de comportement, on est amené à tenir compte du temps d'attente non à l'arrêt où les usagers montent dans un autobus, mais à celui où ils en descendent. Dans le cas des lignes communes, cela pose un problème : un usager qui descend à un arrêt multiple peut être monté à un arrêt qui fait partie de la portion commune des lignes impliquées ou à un arrêt qui n'en fait pas partie; or, le temps d'attente n'étant pas le même dans chaque cas, il devient impossible de déterminer a priori la bonne valeur à utiliser. Dans un souci de simplification, on fait l'hypothèse que tous les usagers qui descendent d'un autobus d'une ligne donnée à un arrêt donné et qui proviennent de la même origine, ont attendu aussi longtemps (la valeur du temps d'attente utilisée étant choisie de manière quelque peu arbitraire), mais cela peut provoquer des distorsions assez importantes dans l'évaluation du coût des itinéraires pour une paire O-D donnée et

conséquemment, à cause de l'hypothèse de comportement, des distorsions dans la répartition des usagers entre les divers itinéraires efficaces disponibles pour cette paire O-D (voir section 3.5). Leyvraz (1980) a souligné plusieurs des problèmes qui peuvent se poser dans l'utilisation de NOPTS.

3.2.8 Modèle de Spiess (1983)

Spiess a été le premier à mettre en lumière la différence très nette qui existe entre un itinéraire et une stratégie du point de vue de l'usager d'un réseau de transport en commun. On peut sans doute dire qu'il a aussi été le premier à définir un véritable modèle d'affectation des passagers pour les réseaux avec arrêts multiples quand la congestion est négligeable.

Après avoir indiqué qu'à la limite, on peut imaginer toutes sortes de stratégies pour les usagers, Spiess se restreint à des stratégies où le choix d'un usager à un arrêt multiple est consituté par un sous-ensemble des lignes qui le desservent. Cependant, contrairement à ce que l'on retrouve dans TRANSCOM, les lignes peuvent diverger par la suite et, même sur les sections communes, l'arrêt auquel descend l'usager n'a pas à être le même pour toutes les lignes.

Spiess montre ensuite que la recherche des stratégies optimales pour tous les usagers se dirigeant vers une destination donnée peut être ramenée à la résolution d'un programme linéaire, à la condition que l'on puisse fournir au modèle les temps d'attente et les répartitions entre les lignes découlant des divers choix possibles des usagers aux arrêts.

Quand les intervalles interautobus suivent des lois exponentielles, il est possible d'exprimer les temps d'attente et les répartitions entre les lignes en fonction de leur fréquence; ceci permet de simplifier le programme linéaire obtenu auparavant et de définir un algorithme efficace pour résoudre le problème.

Pour terminer, soulignons qu'alors que la majorité des algorithmes d'affectation des passagers affectent en même temps tous les usagers d'une origine donnée vers toutes les destinations, celui de Spiess travaille à l'inverse en affectant au même moment tous les usagers se dirigeant vers une destination donnée. Cela est rendu nécessaire par le fait qu'aux arrêts multiples, le cheminement optimal d'un usager dans le réseau comporte des éléments aléatoires. On peut d'ailleurs interpréter la recherche de la stratégie optimale d'un usager comme un programme dynamique probabiliste, or il est bien connu qu'un tel problème doit être résolu par induction à rebours (c'est-à-dire dans l'ordre inverse du déroulement du processus étudié).

3.3 Les modèles d'affectation multimodale

3.3.1 Remarques générales

Ces modèles ont pour but de résoudre de manière globale trois problèmes importants qui, dans la réalité, sont fortement inter-reliés : celui de la répartition modale, celui de l'affectation du trafic routier et celui de l'affectation des passagers.

Ordinairement, pour solutionner ces problèmes, on utilise l'approche traditionnelle (décrite à la section 2.6) qui consiste à effectuer d'abord la répartition modale de la demande totale de transport, puis à affecter de manière indépendante le trafic routier et les usagers du réseau de transport en commun. Ce faisant, on introduit une distorsion dans la répartition modale puisqu'en général, les coûts généralisés découlant des affectations ne correspondent pas exactement à ceux qui ont été utilisés au départ dans le modèle de répartition modale.

Pour résoudre cette difficulté, plusieurs auteurs ont proposé des modèles intégrés de répartition et d'affectation multimodale. On ne peut pas dire que ces modèles soient vraiment satisfaisants, car la grande complexité du problème étudié a amené leurs concepteurs à faire certaines simplifications plus ou moins heureuses.

Dans la description de ces modèles, nous nous en tiendrons aux aspects relatifs à l'affectation des passagers.

3.3.2 TRANSEPT (Daly, 1973; Last et Leak, 1976)

TRANSEPT est fondamentalement un modèle d'affectation des passagers dans lequel on permet aussi aux usagers d'utiliser l'automobile ou la marche pour effectuer leurs déplacements.

L'hypothèse de comportement est une modification de celle proposée par Dial (1971). Pour chaque paire O-D, on détermine les n* meilleurs itinéraires par autobus et on utilise un modèle logistique à (n*+2) dimensions pour faire à la fois la répartition modale (les autres modes

considérés sont l'automobile et la marche) et l'affectation des usagers aux itinéraires.

Dans TRANSEPT, on accorde beaucoup d'importance à la congestion du réseau de transport en commun et, plus particulièrement, à l'impact de celle-ci sur les temps d'attente aux arrêts. Ces temps d'attente sont calculés au moyen d'un modèle de file d'attente qui n'est malheureusement pas décrit de façon détaillée.

Pour les sections de lignes communes, on fait comme Dial et Le Clercq l'hypothèse que les lignes sont équivalentes (en termes de vitesse et de tarification) sur leurs parties communes. Le calcul des temps d'attente et la répartition des usagers entre les lignes sont effectués par une procédure heuristique qui n'est pas décrite dans les articles présentant le modèle.

Il faut souligner que, malgré toutes ses qualités, TRANSEPT est un modèle aux possibilités limitées. Le mécanisme de calcul des temps d'attente aux arrêts exige, en effet, qu'on charge les usagers au fur et à mesure sur les sections de lignes, et qu'au moment où l'on charge une section, toutes les sections "en amont" (c'est-à-dire les sections qui amènent des usagers vers celle que l'on considère à cet instant) aient déjà été chargées. Dans le cas de Coventry (ville pour laquelle TRANSEPT a été mis au point), il était possible d'ordonner correctement les sections de lignes, car la ville possède un réseau de transport à structure radiale; il nous paraît douteux qu'on puisse trouver énormément de villes (surtout des villes nord-américaines) pour lesquelles on puisse faire de même.

3.3.3 Modèle de Florian (1977)

Ce modèle représente une des premières tentatives d'intégrer un modèle d'affectation des passagers à un modèle d'équilibre pour le trafic routier. Le modèle tient compte de la congestion du réseau routier et de l'interaction entre les autobus et les automobilistes sur celui-ci, mais laisse de côté la question de la congestion du système de transport en commun. Dans l'article original, le sous-modèle d'affectation des passagers était très simple : pour chaque paire O-D, les usagers étaient affectés à l'itinéraire à coût minimum, et le problème des lignes communes n'était pas soulevé. Une implantation ultérieure du modèle fut réalisée en utilisant plutôt TRANSCOM comme sous-modèle d'affectation des passagers ; c'est le logiciel EMME (Florian et al., 1979). Ceci fut possible parce que le modèle de Florian est compatible avec n'importe quel sous-modèle de transport en commun dans lequel le coût de transport (en commun) est défini de manière unique pour chaque paire O-D.

3.3.4 Modèle de Schéele (1976)

Pour sa part, Schéele a proposé un modèle d'affectation multimodale pour les cas où la capacité des lignes d'autobus ne peut être négligée. L'hypothèse de comportement retenue est celle, fort discutable, de l'optimisation coopérative. Les effets de la capacité des autobus sont représentés dans le modèle par des contraintes de capacité explicite sur les sections de lignes. Le temps d'attente aux arrêts est égal à la moitié de l'intervalle interautobus moyen de la ligne qui les dessert. Pour ce qui est des lignes communes, le modèle ne semble pas en tenir compte, bien que cela ne soit dit de manière explicite.

3.3.5 Modèle d'Abdulaal et LeBlanc (1979)

Suite aux travaux de Florian, Abdulaal et LeBlanc se sont fixé le but extrêmement ambitieux de développer un modèle d'équilibre global pour la répartition modale, l'affectation du trafic et l'affectation des passagers. Ils ont atteint leur but, mais au prix de simplifications radicales dans la modélisation du réseau de transport en commun. Leur réseau ne comporte que des arcs de transport en commun, sans structure de lignes. La congestion du système de transport en commun n'a pas un effet sur les temps d'attente aux arrêts, mais plutôt sur le coût généralisé des arcs de transport en commun. Bref, le modèle est complètement irréaliste dans sa représentation du transport en commun.

3.4 Les modèles de design de réseaux de transport en commun

3.4.1 Remarques générales

Le design de réseaux de transport en commun est un problème éminemment complexe, car il implique dans sa forme la plus générale la détermination de la localisation des lignes de transport en commun et des fréquences sur celles-ci. La fonction économique à optimiser (implicitement ou explicitement) est, elle aussi, complexe, car elle doit permettre de réaliser un arbitrage entre les coûts d'exploitation du système et la qualité du service offert aux usagers. C'est d'ailleurs pour mesurer adéquatement la qualité du service offert qu'il faut réaliser une affectation des passagers dans les modèles de design.

3.4.2 Etude de Lampkin et Saalmans (1967)

Le but de l'étude de Lampkin et Saalmans était la réorganisation complète du système de transport en commun d'une petite ville d'Angleterre. Devant l'impossibilité d'arriver à un modèle global satisfaisant, il fut décidé de séparer le problème en plusieurs parties dont une impliquait l'affectation des passagers sur les réseaux proposés.

Au niveau de l'affectation elle-même, l'hypothèse de comportement est celle de l'optimisation égoïste et les problèmes de capacité des autobus sont ignorés (on procède cependant à des ajustements de fréquences sur les lignes dans le modèle d'allocation optimale des autobus pour s'assurer d'une capacité suffisante sur chaque ligne). On suppose, d'autre part, que les intervalles entre les autobus d'une ligne sont fixes et, en conséquence, le temps d'attente à un arrêt donné est égal à la moitié de l'intervalle interautobus. Pour les arrêts multiples, on définit a priori pour chaque paire O-D un sous-ensemble de lignes "efficaces" et on en déduit le temps d'attente et la répartition des usagers entre les lignes en utilisant les hypothèses faites sur les distributions des intervalles entre les autobus d'une même ligne (lois déterministes).

3.4.3 Modèle de Holm (1973)

Holm ne résout pas vraiment le problème de design de réseau, mais plutôt celui de l'allocation optimale d'une flotte d'autobus pour une structure de lignes donnée. L'hypothèse de comportement qu'il suggère est curieuse : pour chaque paire O-D, les usagers sont répartis entre les itinéraires disponibles inversement proportionnellement au coût

généralisé de chacun. Les temps d'attente sont donnés par une courbe empirique basée sur les travaux de Holroyd et Scraggs (1966). Il n'y a pas de lignes communes dans ce modèle.

3.4.4 Modèle de Sharp (1973)

Sharp a sans doute été le premier à proposer un modèle global pour le design de réseaux de transport en commun. Dans sa formulation du problème, Sharp a dû ramener dans une fonction économique unique les coûts d'exploitation du réseau et les coûts généralisés de déplacement des usagers. Cela revient implicitement à supposer que les usagers se comportent selon la règle de l'optimisation coopérative. La capacité limitée des autobus est modélisée par des contraintes de capacité explicites sur les sections de lignes. Le temps d'attente aux arrêts est égal à la moitié de l'intervalle interautobus moyen. Ce modèle ne permet pas les lignes communes.

3.4.5 Modèle de Mandl (1980)

Mandl a suggéré de séparer le problème de design en trois sous-problèmes (affectation des passagers, allocation des autobus aux lignes et localisation des lignes) qui peuvent être résolus itérativement jusqu'à l'obtention d'une solution satisfaisante. Son sous-modèle d'affectation des passagers, basé en partie sur les travaux de Chriqui et Robillard, ressemble beaucoup à TRANSCOM (voir section 3.2.5). Les deux modèles diffèrent principalement dans certains aspects du traitement des lignes communes :

a) Mandl utilise une représentation explicite pour les différentes

combinaisons de lignes communes (ce qui peut causer des difficultés s'il y a beaucoup de lignes communes).

b) Mandl n'utilise pas l'algorithme de choix du sous-ensemble optimal de lignes communes.

c) Le modèle de Mandl ne semble pas être en mesure de traiter le cas où des lignes communes ont des vitesses très différentes.

3.4.6 Approche de Hasselström (1981)

Dans sa thèse de doctorat, Hasselström propose une approche par raffinements successifs pour le problème de design de réseaux de transport en commun. Cela l'amène à développer une série de modèles de plus en plus détaillés. Associés à ces modèles, on retrouve plusieurs sous-modèles d'affectation des passagers. Le plus détaillé d'entre eux repose sur l'hypothèse d'une exploitation à horaires, ce qui lui enlève beaucoup d'intérêt à notre point de vue. Le phénomène de congestion n'entre pas directement en ligne de compte dans ce sous-modèle (c'est une autre partie du modèle qui fait les ajustements nécessaires aux fréquences des lignes pour assurer un niveau de service suffisant). Pour déterminer les temps d'attente et les répartitions des usagers entre lignes aux arrêts multiples, on utilise une approximation heuristique basée sur une idée de réduction de fréquence pour les lignes moins attrayantes (lignes moins rapides). Selon Hasselström, cette approximation a donné d'excellents résultats en pratique.

Notons, en terminant, que dans sa thèse, Hasselström reprend plusieurs résultats qui sont le fruit de recherches antérieures effectuées en compagnie d'Andréasson (1979).

3.5 Commentaires sur les hypothèses de comportement

Dans tout modèle visant à représenter le comportement humain, l'hypothèse fondamentale que l'on fait à ce sujet est cruciale. Il nous apparaît donc essentiel de nous attarder quelque peu sur les hypothèses de comportement utilisées dans les modèles que nous venons de présenter.

Considérons d'abord les modèles qui utilisent l'hypothèse d'optimisation coopérative. Cette hypothèse nous a toujours paru être un choix totalement inadéquat pour un modèle d'affectation des passagers. Certains auteurs, dont Schéele (1976, page 2), ont essayé de justifier ce choix en faisant remarquer que les systèmes de transport en commun étant sous le contrôle d'autorités centrales, il était naturel dans ce cas de vouloir atteindre une solution qui optimise le bien-être de l'ensemble des usagers. L'argument est fallacieux. En effet, bien qu'il nous paraisse tout à fait normal et désirable que la direction d'un système de transport en commun se fixe comme but de fournir aux usagers le meilleur service possible pour l'ensemble d'entre eux, il nous semble utopique de croire que chaque usager a des préoccupations "collectives" quant il choisit sa stratégie de transport personnelle. Le véritable défi pour le planificateur d'un réseau de transport en commun consiste, d'après nous, à réaliser le meilleur système possible d'un point de vue global, compte tenu du comportement individualiste des usagers.

Comme nous l'avons indiqué auparavant à la sous-section 3.2.1, s'il n'y a pas dans un modèle d'interaction entre les usagers (que ce

soit par des contraintes de capacité ou par des coûts de parcours qui varient en fonction du nombre d'usagers), les hypothèses d'optimisation égoïste et d'optimisation coopérative sont équivalentes. Il nous paraît préférable, suite à la discussion qui précède, d'analyser ces situations comme des situations d'optimisation égoïste; nous y reviendrons donc un peu plus loin.

Reste le cas des modèles où l'on a l'hypothèse d'optimisation coopérative et une interaction des usagers (Sharp, Schéele). Intuitivement, on peut penser que ce genre de modèles va produire des affectations irréalistes, entre autres, des affectations dans lesquelles certains passagers doivent descendre d'un autobus qui les amènerait rapidement jusqu'à leur destination, pour laisser "la place" à d'autres usagers. Dans les faits, il est très facile de construire des exemples simples, mais plausibles, qui permettent de vérifier cette intuition (voir Gendreau, 1977, pages 16 à 20).

L'hypothèse d'optimisation égoïste est, à première vue, extrêmement séduisante, ce qui explique sans doute que la grande majorité des modèles d'affectation des passagers l'aient adoptée implicitement ou explicitement. Néanmoins, son utilisation pose certains problèmes en pratique.

Pour les modèles qui négligent complètement l'interaction entre les usagers au niveau de l'affectation proprement dite (Dial, Fearnside et Draper, Le Clercq, TRANSCOM, Spiess, Florian, Lampkin et Saalmans, Mandl), cette hypothèse implique que pour chaque paire O-D, les usagers utilisent le même itinéraire ou la même stratégie (dans le cas où on

travaille avec des itinéraires, on obtient alors ce qu'on appelle des affectations "tout ou rien" ou des affectations "tout ou rien" avec diversion sur les lignes communes). Or, il est clair qu'en général les usagers allant d'une origine vers une destination n'utilisent pas tous exactement le même itinéraire (stratégie), et ce pour diverses raisons : perception différente des composantes du déplacement, connaissance imparfaite du réseau de transport, etc. Qui plus est, dans le cas où pour une paire O-D, plusieurs "bons" itinéraires sont disponibles, une modification même très légère de certains coûts de transport peut provoquer un changement total de l'affectation produite.

L'hypothèse de Dial (et ses proches parentes comme l'hypothèse utilisée dans TRANSEPT) représente une tentative de solution des problèmes soulevés relativement aux affectations "tout ou rien" puisque pour chaque paire O-D, elle "disperse" les usagers sur les différents itinéraires "efficaces" disponibles. Il semble cependant que l'application de cette hypothèse de comportement comporte aussi sa large part de difficultés, et que les affectations obtenues soient peu réalistes (voir à ce sujet Burrell, 1976; Florian et Fox, 1976; Sheffi et Daganzo, 1978; Leyvraz, 1980; Florian, 1981).

Suite à cette discussion, il devient clair qu'il n'est pas facile d'énoncer une hypothèse de comportement qui soit à la fois simple et adéquate. Personnellement, nous persistons à penser que l'hypothèse d'optimisation égoïste est la plus valable qui ait été proposée à ce jour si elle est utilisée dans un modèle dans lequel le coût des itinéraires croît avec le nombre d'usagers qui les empruntent (c'est-à-dire

un modèle d'équilibre). On sait qu'alors les usagers d'une paire O-D ne sont pas tous affectés au même itinéraire, sauf cas exceptionnel. De plus, certaines études empiriques (Florian et Nguyen, 1976a) ont montré que les modèles d'équilibre pour l'affectation du trafic routier produisent des résultats assez proches de ce que l'on peut observer dans la réalité.

3.6 Itinéraires versus stratégies

Nous avons insisté dans la sous-section 3.2.1 sur le fait que les modèles d'affectation des passagers ont presque tous été définis en termes du choix d'itinéraire, et non de stratégie, qu'effectuent les usagers. Spiess (1983, page 2) fait remarquer très justement qu'alors on ne considère qu'un sous-ensemble sous-optimal des possibilités de cheminement dans le réseau.

Pour illustrer ceci, nous avons choisi un exemple très simple de réseau de transport en commun. Cet exemple nous permettra aussi de voir la différence qui existe entre les algorithmes qui cherchent un "itinéraire simple" (suite de sections de lignes) comme celui de Fearnside et Draper, et ceux qui cherchent un "itinéraire avec diversion" (séquence de points de correspondance avec diversion possible sur les lignes communes) comme ceux de Dial, Le Clercq, TRANSCOM et Mandl.

Soulignons que nous avons choisi d'utiliser les mêmes temps d'attente aux arrêts pour tous les algorithmes, même si certains des modèles considérés diffèrent sur cet aspect. Pour les arrêts multiples, nous appliquons la règle de combinaison des fréquences. D'autre part, pour les

sections de lignes communes, nous avons supposé que les vitesses des lignes étaient égales, ce qui permet d'appliquer les algorithmes de Dial, Le Clercq et Mandl sans problème.

Les résultats sont présentés en distinguant les solutions obtenues par l'application des algorithmes de recherche d'itinéraires simples (ARIS : Fearnside et Draper), des algorithmes de recherche d'itinéraires avec diversion (ARID : Dial, Le Clercq, TRANSCOM, Mandl) et des algorithmes de recherche de stratégies optimales (ARSO : Spiess).

Exemple 3.1

Supposons que l'on ait le réseau suivant

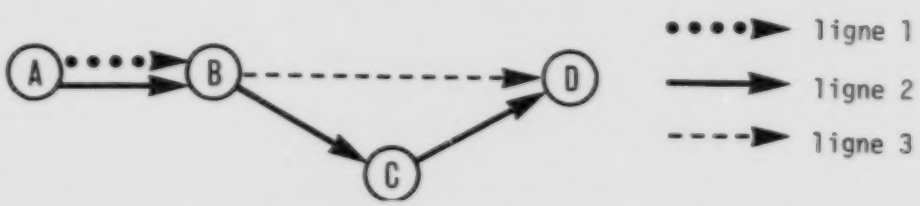


Figure 2 : Réseau de l'exemple 3.1.

et que l'on veuille déterminer le cheminement optimal d'un usager se dirigeant de A vers D. Le temps de parcours est de 8 mn pour chaque section de ligne et les temps d'attente pour les différentes combinaisons de lignes sont donnés par le tableau suivant :

Lignes	Temps d'attente
1	12 mn
2	12
3	12
1 et 2	6
2 et 3	6

Si l'on suppose que les composantes du déplacement (temps attente, temps à bord des autobus) ont la même pondération, on obtient les solutions "optimales" qui suivent.

Solution ARIS

Prendre la ligne 2 en A et continuer jusqu'à D.

Solution ARID

Prendre la ligne 1 ou la ligne 2 en A, descendre en B et prendre la ligne 3 jusqu'à D.

Solution ARSO

Prendre la ligne 1 ou la ligne 2 en A; si, en A, un autobus de la ligne 1 est pris, descendre en B et prendre la ligne 2 ou la ligne 3 jusqu'à D; si, en A, un autobus de la ligne 2 est pris, continuer jusqu'à D.

Le tableau suivant indique la probabilité d'emprunter chacune des sections de lignes selon l'algorithme utilisé.

Ligne	Section	ARIS	ARID	ARSO
1	A → B	0	1/2	1/2
2	A → B	1	1/2	1/2
2	B → C	1	0	3/4
2	C → D	1	0	3/4
3	B → D	0	1	1/4

Pour chaque solution, les espérances du temps d'attente (E [TA]), du temps passé à bord d'autobus (E [TB]) et du temps total de parcours (E [TT]) sont donnés par le tableau qui suit.

Algorithme	E [TA]	E [TB]	E [TT]
ARIS	12mn	24 mn	36 mn
ARID	18	16	34
ARSO	9	22	31

Comme on le voit, l'espérance du temps total de parcours décroît nettement quand on considère des stratégies plus complexes pour les usagers.

Avant de laisser cet exemple, il faut remarquer que si le temps de parcours de la section A → B par la ligne 1 était ramené à 6 mn, on ne pourrait pas appliquer directement les algorithmes de Dial, Le Clercq et Mandl; il faudrait soit décider arbitrairement d'une valeur quelconque pour le temps de parcours de la section A → B (entre les 6 mn de la ligne 1 et les 8 mn de la ligne 2), soit utiliser un artifice de modélisation pour représenter les lignes 1 et 2 comme totalement distinctes. Dans un cas comme dans l'autre, on aboutit à une solution qui n'est pas totalement satisfaisante. Si, d'autre part, on utilisait TRANSCOM ou l'algorithme de Spiess, on pourrait déterminer sans aucune difficulté la solution du problème modifié (qui est identique à celle du problème original, mais avec une réduction dans chaque cas de 1 mn pour E [TA] et E [TT]).

3.7 Critiques relatives au traitement de la congestion

Dans l'introduction, nous avons insisté sur le fait que la congestion du système de transport en commun causée par la capacité limitée des autobus pouvait avoir un impact important sur les temps d'attente

des usagers et, en conséquence, sur l'affectation de ceux-ci sur le réseau. Cependant, en regard des seize modèles que nous venons de passer en revue, on pourrait être amené à mettre en doute cette assertion. En effet, dans douze de ces modèles (Dial, Fearnside et Draper, Le Clercq, TRANSCOM, Andréasson, NOPTS, Spiess, Florian, Lampkin et Saalmans, Holm, Mandl et Hasselström), le phénomène n'a aucun impact direct sur l'affectation, soit qu'il soit tenu pour négligeable, soit qu'on en tienne compte de manière exogène par des ajustements de fréquences a posteriori. Des quatre modèles restants, un est complètement irréaliste (celui d'Abdulaal et LeBlanc) et deux autres sont des modèles d'optimisation coopérative avec contraintes de capacité explicite sur les sections de lignes (Schéele et Sharp). C'est donc dire que sur les quinze modèles qui nous restent après avoir écarté celui d'Abdulaal et LeBlanc, il y en a quatorze dans lesquels la congestion n'a aucun impact si la capacité est tout juste suffisante pour que l'on puisse transporter tous les usagers.

Pourtant, l'étude des files d'attente nous apprend que même avec un taux d'achalandage relativement faible (de l'ordre de 0,5), les temps d'attente deviennent assez sensibles aux variations dans la charge du système. D'autre part, nous avons mentionné précédemment qu'il suffit souvent dans la plupart des modèles d'une variation très légère de certains coûts de déplacement pour provoquer des changements importants dans l'affectation produite. C'est donc dire que pour être certain que la congestion a un impact négligeable sur l'affectation, il faut que l'achalandage du système de transport en commun soit très faible. Nous nous permettons de douter que ce soit en général le cas quand on s'intéresse au comportement du réseau pendant les heures de pointe.

Jusqu'à maintenant, nous n'avons pas parlé de TRANSEPT. A la lecture des deux articles décrivant le modèle (en particulier, celui de Last et Leak), il est évident que ses concepteurs avaient perçu très clairement l'impact de la congestion sur les temps d'attente et la répartition des usagers entre les lignes communes. On ne peut que regretter le fait que TRANSEPT ne soit applicable qu'aux réseaux de transport à structure radiale et la très grande concision de ses auteurs qui ne décrivent pas en détail les sous-modèles d'attente et de répartition entre les lignes communes.

3.8 Remarques sur le calcul des temps d'attente

Le but de cette section est d'examiner la qualité des fonctions d'attente utilisées dans les modèles décrits auparavant si l'on fait abstraction des effets de la congestion sur les temps d'attente.

3.8.1 Temps d'attente aux arrêts simples

Au niveau des arrêts simples, on peut classifier les modèles présentés en quatre catégories :

a) les modèles qui n'établissent pas de lien explicite entre les temps d'attente et les intervalles interautobus des lignes impliquées;

b) les modèles qui posent le temps d'attente égal à la moitié de l'intervalle interautobus moyen;

c) les modèles qui postulent des distributions exponentielles pour les intervalles interautobus;

d) les modèles qui utilisent des courbes empiriques pour les temps d'attente.

On peut placer le modèle de Abdulaal et LeBlanc dans la catégorie a). Il doit être clair que la modélisation du phénomène d'attente aux arrêts n'était pas dans ce cas une préoccupation première, ce qui explique le peu d'intérêt apporté à la question.

Dans la catégorie b), on retrouve plusieurs modèles, à savoir ceux de Dial, Fearnside et Draper, Le Clercq, Andréasson, Schéele, Lampkin et Saalmans, Sharp et Hasselström et NOPTS. Holroyd et Scraggs (1966) ont montré que si les usagers se présentent au hasard à un arrêt et s'ils peuvent monter à bord du premier autobus qui s'y présente, alors le temps d'attente moyen à cet arrêt (W) est donné par la formule

$$W = \frac{1}{2}\left(T + \frac{\sigma^2}{T}\right)$$

où T est l'intervalle interautobus moyen

et σ^2 est la variance de l'intervalle interautobus.

C'est donc dire que dans les modèles de la catégorie b), on suppose que $\sigma^2 = 0$, c'est-à-dire que pour chaque ligne, les intervalles interautobus sont fixes et égaux entre eux. Il est quelque peu agaçant de devoir noter que seuls Lampkin et Saalmans ont cru bon de préciser explicitement qu'ils faisaient cette hypothèse. Nous discuterons plus à fond des distributions choisies pour les intervalles interautobus à la sous-section 3.8.3.

TRANSCOM, le modèle de Mandl et la version simplifiée du modèle de Spiess utilisent des lois exponentielles pour les intervalles interautobus,

car il est alors très facile de calculer les temps d'attente moyens aux arrêts (ils sont égaux aux intervalles interautobus moyens). On peut cependant se demander si ce choix de distribution est très approprié (voir sous-section 3.8.3). Soulignons ici que dans sa version générale, le modèle de Spiess peut être utilisé avec une distribution quelconque, mais c'est alors à l'utilisateur du modèle que revient la tâche de déterminer les temps d'attente aux arrêts.

Plutôt que de s'en remettre à des dérivations théoriques pour les temps d'attente, Holm et les concepteurs de TRANSEPT ont préféré travailler avec des courbes empiriques. Holm a extrapolé la sienne des résultats de l'étude d'Holroyd et Scraggs (1966) menée à Londres, tandis que pour TRANSEPT, on a réalisé une étude spécifique à Coventry (ville-cible du modèle). Dans les deux cas, on est arrivé à la conclusion que le temps d'attente moyen est inférieur à la moitié de l'intervalle interautobus moyen, ce qui semblerait indiquer que les usagers ne se présentent pas aux arrêts au hasard!

Cette méthodologie est intéressante, mais comme le fait remarquer Daly (1973, page 7), les usagers de villes différentes ont probablement un comportement différent, ce qui veut dire que son application dans une nouvelle ville est conditionnelle à la réalisation d'études du genre de celles menées à Londres et à Coventry.

3.8.2 Temps d'attente aux arrêts multiples

Pour les arrêts multiples, on retrouve encore plusieurs approches pour le calcul des temps d'attente (et des répartitions des usagers entre les lignes).

Il y a d'abord celle de TRANSEPT qui procède par une méthode de convolutions de distributions, et dont on peut supposer qu'elle est adéquate, sans toutefois être en mesure de l'affirmer, car elle n'est qu'évoquée brièvement dans l'article de Last et Leak.

Pour les modèles utilisant des lois exponentielles pour les intervalles interautobus (TRANSCOM, Mandl, version simplifiée de Spiess), le calcul des temps d'attente et des répartitions entre les lignes ne pose aucun problème. Comme l'ont montré Chriqui et Robillard (1975), le temps d'attente est égal à l'inverse de la fréquence combinée des lignes considérées et la probabilité de prendre un autobus d'une ligne donnée est proportionnelle à la fréquence de celle-ci.

Parmi les modèles dans lesquels on suppose les intervalles interautobus d'une même ligne fixes et égaux, ceux de Dial et Le Clercq et NOPTS appliquent la règle de la fréquence combinée et celle de la répartition proportionnelle aux fréquences. Dans ce cas-ci, c'est une approximation grossière qui peut causer des erreurs relativement importantes comme nous le verrons dans l'exemple 3.2.

Lampkin et Saalmans, de leur côté, ont décidé de calculer directement les valeurs désirées par un raisonnement probabiliste. Pour chaque ligne passant à un arrêt donné, ils définissent implicitement une variable aléatoire correspondant à la longueur de l'intervalle de temps entre l'arrivée (au hasard) d'un usager à l'arrêt et le passage du prochain autobus de cette ligne. Il suffit ensuite pour déterminer le temps d'attente moyen à l'arrêt de calculer l'espérance du minimum des variables aléatoires

associées aux lignes. Bien qu'en apparence inattaquable, le raisonnement de Lampkin et Saalmans comporte une faille de taille. En effet, dans leur calcul du minimum des variables aléatoires associées aux lignes, ils font implicitement l'hypothèse que ce sont des variables indépendantes, or elles ne le sont pas nécessairement! On peut expliquer ceci par le caractère périodique que prend le processus de passage des autobus à l'arrêt quand les intervalles interautobus suivent des lois déterministes. Dans un tel cas, il existe une certaine synchronisation entre les lignes qui fait que les passages des autobus des diverses lignes ne sont pas des événements indépendants. L'exemple 3.2 permettra de voir ceci de manière plus claire.

Dans la version générale de son modèle, Spiess reprend certains résultats de la théorie du renouvellement pour obtenir des expressions (pas nécessairement calculables en pratique) du temps d'attente et des probabilités de répartition entre les lignes. Il est remarquable de noter qu'il tombe lui aussi dans le même piège que Lampkin et Saalmans en postulant a priori l'indépendance des temps d'attente des diverses lignes.

Nous allons maintenant présenter un exemple qui permettra de voir que la règle de la fréquence combinée ne tient pas pour des temps inter-autobus déterministes et que l'hypothèse d'indépendance entre les temps d'attente des lignes n'est pas vérifiée.

Exemple 3.2

Supposons qu'à un arrêt donné, il passe deux lignes d'autobus dont les fréquences sont égales et que pour chaque ligne, les intervalles inter-autobus sont fixes et égaux entre eux. Supposons aussi, sans perte de généralité, que les deux lignes ont fréquence 1 (et donc intervalles inter-autobus égaux à 1 aussi) et que le "déphasage" de la ligne 2 par rapport à la ligne 1 est égal à δ $(0 < \delta < 1)$. Pour simplifier la suite de la discussion, fixons arbitrairement le passage du premier autobus à 0. C'est donc dire que les passages des autobus de la ligne 1 se produisent aux instants 0, 1, 2, ... et ceux des autobus de la ligne 2 aux instants δ, $1+\delta$, $2+\delta$, ...

La règle des fréquences combinées pour les lois déterministes nous dit que le temps d'attente moyen à cet arrêt est égal à la moitié de l'inverse de la fréquence combinée. Appelons W_{FC} cette valeur. On a donc $W_{FC} = \frac{1}{4}$. D'autre part, les fréquences étant égales, on a par la règle de répartition proportionnelle aux fréquences $p_1^{RP} = p_2^{RP} = \frac{1}{2}$.

Appliquons maintenant l'approche de Lampkin et Saalmans ou celle de Spiess à cet exemple. Dans un cas comme dans l'autre, on trouve que l'intervalle entre l'arrivée d'un usager et le passage du prochain autobus suit une loi uniforme sur (0,1) pour chacune des deux lignes. On peut montrer facilement que l'espérance du minimum de deux variables aléatoires uniformes (0,1) indépendantes est égal à $\frac{1}{3}$. On pose donc $W_{LS} = \frac{1}{3}$. D'autre part, comme on a la même loi pour le temps d'attente pour chaque ligne, on obtient directement par un argument de symétrie les probabilités de répartition $p_1^{LS} = p_2^{LS} = \frac{1}{2}$.

Livrons-nous maintenant à une analyse exacte de ce qui se passe à cet arrêt. Considérons le cas d'un usager qui arrive au hasard. Soit $(k,k+1]$ l'intervalle dans lequel il arrive. Maintenant, il y a deux cas possibles :

a) Il arrive dans le sous-intervalle $(k,k+\delta]$. La probabilité que ceci se passe est δ. Il prend alors la ligne 2 et il attend en moyenne $\frac{\delta}{2}$.

b) Il arrive dans le sous-intervalle $(k+\delta, k+1]$. La probabilité que ceci se passe est $(1-\delta)$. Il prend alors la ligne 1 et il attend en moyenne $\frac{1-\delta}{2}$.

On peut tirer de ce qui précède que les probabilités de répartition exactes sont $p_1 = (1-\delta)$ et $p_2 = \delta$. Pour l'espérance globale du temps d'attente, on obtient en conditionnant sur la ligne prise

$$W = \frac{(1-\delta)^2}{2} + \frac{\delta^2}{2} \quad .$$

On peut donc voir que le déphasage qui est négligé dans tous les modèles a un impact déterminant sur la répartition des usagers et sur le temps d'attente. On peut d'ailleurs noter que

. $\lim_{\delta \to 0} W = \lim_{\delta \to 1} W = \frac{1}{2}$

. $W = \frac{1}{4} \iff \delta = \frac{1}{2}$.

C'est donc dire que les résultats de Dial et Le Clercq correspondent à l'hypothèse de déphasage optimal ($\delta = \frac{1}{2}$). Quant aux résultats de

Lampkin et Saalmans, il n'existe aucune valeur de δ pour laquelle $W = \frac{1}{3}$ et $p_1 = p_2 = \frac{1}{2}$ (puisque $p_1 = p_2 = \frac{1}{2} \Rightarrow \delta = \frac{1}{2} \Rightarrow W = \frac{1}{4}$). Fin de l'exemple.

Avant de clore cette discussion, nous aimerions faire une dernière remarque au sujet de la règle des fréquences combinées dans le cas déterministe. Si l'on considère tous les passages d'autobus à un arrêt donné de manière globale, on obtient une nouvelle série d'intervalles interautobus qui a une certaine moyenne $\tilde{T}$ et une certaine variance $\tilde{\sigma}^2$. Il est facile de voir que $\tilde{T}$ est égal à l'inverse de la fréquence combinée des lignes. Maintenant appliquons le résultat de Holroyd et Scraggs à ces nouveaux intervalles. On obtient

$$W = \frac{1}{2}\left(\tilde{T} + \frac{\sigma^2}{\tilde{T}}\right) \quad .$$

C'est donc dire que Dial et Le Clercq font sans s'en rendre compte l'hypothèse que $\tilde{\sigma}^2 = 0$. Il est assez facile de se rendre compte que le seul cas où l'on peut avoir $\tilde{\sigma}^2 = 0$ est celui où toutes les fréquences sont égales et que les déphasages sont optimaux.

3.8.3 Choix des distributions des intervalles interautobus

Pour des raisons évidentes de simplicité, les auteurs des modèles que nous avons présentés se sont bornés à n'utiliser que des distributions déterministes ou exponentielles pour les intervalles interautobus. Cela nous semble fort regrettable, car ces deux distributions représentent, chacune à leur façon, des cas extrêmes. La loi déterministe correspond à un service d'autobus d'une régularité et d'une fiabilité

infaillibles, tandis que la loi exponentielle (avec un écart-type égal à la moyenne) implique un service aléatoire sans mémoire. Il nous semble certain qu'il y a toujours dans un service d'autobus une certaine variabilité (surtout aux heures de pointe quand la circulation est dense), mais il serait désespérant qu'elle soit aussi grande que celle que l'on obtient quand on utilise des lois exponentielles. Qui plus est, dans tous les modèles développés jusqu'à maintenant (à l'exception de la version générale de celui de Spiess), les intervalles interautobus suivent la même distribution (pas avec les mêmes paramètres) pour toutes les lignes. Là aussi, il y aurait place pour une nette amélioration. Un modèle adéquat devrait permettre de spécifier un niveau de variabilité différent pour des lignes différentes. Ceci nous amène a proposer d'utiliser la distribution d'Erlang pour les intervalles interautobus. Contrairement à l'exponentielle et à la loi déterministe qui n'ont qu'un seul paramètre (la moyenne), l'Erlang en a deux (la moyenne et le paramètre de forme), ce qui permet de faire varier le coefficient de variation. D'autre part, c'est une distribution assez simple du point de vue mathématique pour que l'on puisse réussir à obtenir des résultats intéressants et applicables. C'est ce que nous montrerons dans les chapitres cinquième et septième.

3.8.4 Le hasard et les arrivées des usagers aux arrêts

Si l'on veut utiliser un modèle théorique pour le calcul des temps d'attente, il est presque certain que l'on a besoin de faire l'hypothèse que les arrivées des usagers aux arrêts se font au hasard, c'est-à-dire, entre autres choses, que les usagers ne choisissent le moment auquel ils

se présentent à l'arrêt en fonction du moment auquel ils prévoient que l'autobus va passer. Pourtant, l'étude de Holroyd et Scraggs et celle menée dans le cadre de TRANSEPT semblent indiquer le contraire. La question se pose donc avec beaucoup d'acuité : l'hypothèse des arrivées au hasard est-elle raisonnable, ou bien induit-elle une distorsion importante dans les résultats produits par les modèles qui reposent sur elle? Il ne semble pas y avoir de réponse universelle à cette question, dans le sens que la réponse dépend de la nature du service offert dans une ville donnée (service fiable ou régulier ou non, horaires fixes et connus ou non) et de certains aspects de la psychologie des usagers. Il faut d'ailleurs noter ici qu'à Coventry où fut réalisée l'étude pour TRANSEPT, les horaires étaient affichés aux arrêts, et qu'à Londres où fut menée celle d'Holroyd et Scraggs, le service était fiable. Il doit cependant être clair que dans le cas d'une ville où le service ne suit pas un horaire très strict (ce qui est probablement le cas de la plupart des villes nord-américaines), l'hypothèse des arrivées au hasard devrait être vérifiée, car les usagers n'ont alors aucun point de repère leur permettant de choisir le moment auquel ils doivent se présenter à l'arrêt.

En terminant, nous croyons bon de rappeler encore une fois que la question fondamentale n'est pas de savoir si l'hypothèse est vérifiée ou non, mais plutôt de déterminer si l'on introduit une distorsion appréciable dans les résultats du modèle en la faisant.

3.9 Conclusion

Au terme de cette revue critique des modèles d'affectation des passagers, il nous paraît important de porter un jugement global sur le travail qui a été fait sur le sujet jusqu'à maintenant. D'une certaine façon, on a l'impression que le problème d'affectation des passagers n'a pas suscité un grand intérêt. Il y a relativement peu de modèles et parmi ceux-ci on sent bien qu'il y en a plusieurs dans lesquels l'affectation des passagers n'a pas semblé retenir très longtemps l'attention des auteurs. Ce qui est peut-être encore plus surprenant, c'est de devoir se rendre compte que dans tous les modèles, il y a au moins un aspect important du problème qui est négligé ou mal modélisé, quand ce n'est pas plusieurs. Il y a bien sûr quelques modèles qui méritent qu'on s'y arrête. Nous pensons, entre autres, à TRANSEPT et au modèle de Spiess. Si l'on pouvait vraiment négliger les effets de la congestion, la version générale du modèle de Spiess serait très satisfaisante. Il faut rendre crédit à Spiess pour sa formulation du problème en termes de stratégies, car il a ainsi clarifié un long malentendu. Dans notre travail sur le problème, nous étions arrivés de notre côté à une formulation similaire mais qui tenait compte de la congestion (c'est d'ailleurs cette formulation que nous présenterons au chapitre sixième, elle ne doit rien au travail de Spiess). Nous croyons que c'est une formulation naturelle si l'on regarde le problème pour lui-même, c'est-à-dire sans se laisser distraire par les résultats obtenus pour le problème d'affectation du trafic routier qui est, au fond, bien différent de celui qui nous intéresse.

Pour ce qui est de TRANSEPT, comment ne pas souligner la manière dont on y traite les effets de la capacité des autobus? Comment ne pas souligner les commentaires fort judicieux de Daly, Last et Leak sur certains aspects du problème? On peut se demander comment il se fait que TRANSEPT n'ait pas été le point de départ d'une série de modèles d'affectation des passagers. Nous croyons que cela est dû en bonne part au fait que TRANSEPT n'est pas un modèle général, mais un modèle adapté à une ville particulière, et aussi, pour une autre part, à la trop grande discrétion de ses auteurs.

A la lumière de tout ce qui a été dit auparavant, nous croyons que deux objectifs s'imposent à ceux qui s'intéressent au problème de l'affectation des passagers :

1) formulation d'un modèle adéquat tenant compte des effets de la congestion;

2) obtention de fonctions d'attente aux arrêts plus réalistes et plus flexibles.

C'est à ces deux problèmes que nous nous attaquons dans la suite de cette thèse.

CHAPITRE QUATRIÈME

FORMULATION DU MODÈLE D'AFFECTATION D'ÉQUILIBRE POUR LES RÉSEAUX SANS ARRÊTS MULTIPLES

4.1 Introduction

Dans le chapitre précédent, nous avons vu que les divers modèles proposés jusqu'à maintenant pour le problème d'affectation des passagers laissaient fort à désirer quant au traitement du phénomène de congestion du réseau de transport en commun et quant à la modélisation du phénomène d'attente aux arrêts. Dans le cadre du présent chapitre, nous développons un modèle d'équilibre d'un flot de plusieurs produits pour le problème d'affectation des passagers avec demande fixe sous l'hypothèse d'optimisation égoïste. Ce modèle est un proche parent des modèles d'équilibre pour l'affectation du trafic routier qui sont maintenant parmi les plus utilisés pour ce dernier problème (voir Florian et Nguyen, 1976b ou Dow et Van Vliet, 1979 pour des exemples de validation de ces modèles).

Il faut remarquer que,dans ce chapitre-ci, nous ne considérons que les réseaux de transport qui ne comportent pas d'arrêts multiples (le cas des réseaux avec arrêts multiples sera étudié au chapitre sixième). De plus, nous ne faisons qu'effleurer ici la question du calcul du temps d'attente aux arrêts,qui sera traitée au chapitre cinquième.

Soulignons finalement que le modèle que nous présentons dans ce chapitre est une version quelque peu modifiée du modèle que nous avions défini auparavant dans notre mémoire de maîtrise (Gendreau, 1977).

4.2 Particularités du modèle

4.2.1 Modèle d'équilibre

Le modèle présenté est un modèle d'équilibre, c'est-à-dire que nous essayons de déterminer un état dit d'équilibre vers lequel va tendre le système étudié, plutôt qu'une solution optimale à un problème donné. Cette notion d'équilibre est à rapprocher de celle qui est utilisée traditionnellement en économie (par exemple, quand on parle d'équilibre entre l'offre et la demande). Elle repose d'ailleurs sur les mêmes hypothèses fondamentales : optimisation par les individus (ou les firmes) de leur bien-être propre (optimisation égoïste), comportement rationnel, connaissance parfaite de la situation (dans notre cas, du réseau de transport et des déplacements des autres usagers), etc. Dans notre modèle, l'équilibre sera atteint quand les flots d'usagers sur le réseau (et les coûts de transport qui en découlent) satisferont le premier principe de Wardrop.

Une remarque relativement au principe de Wardrop et à l'hypothèse d'optimisation égoïste : comme nous ne nous intéressons ici qu'aux réseaux sans arrêts multiples, les stratégies des usagers se ramènent tout simplement à des itinéraires.

4.2.2 Déplacements à pied intégrés

Dans tout modèle d'affectation des passagers, on doit permettre aux usagers d'effectuer certains déplacements à pied, ne serait-ce que pour accéder au réseau de transport en commun. Dans beaucoup de modèles, on

limite ces déplacements à pied autant que possible, forçant, par le fait même, les usagers à utiliser le transport en commun. Pour notre part, nous avons choisi d'intégrer complètement le mode "marche". Nous n'utilisons pas de sous-modèle de répartition modale transport en commun-marche, mais, pour chaque paire origine-destination, il y a un itinéraire qui correspond à un déplacement totalement effectué à pied. Cet itinéraire sera utilisé s'il est plus (ou aussi) avantageux que les autres itinéraires disponibles pour cette paire O-D.

4.2.3 Réseau de transport en commun

Aux fins du modèle, nous avons besoin de connaître deux ensembles d'informations relativement au réseau de transport en commun :

a) sa structure de base sous forme de graphe de base;
b) divers paramètres de service pour les lignes.

Pour construire le graphe de base du réseau, on utilise la procédure décrite dans la section 2.5. Il faut cependant remarquer que, puisque le réseau ne comporte pas d'arrêts multiples, chaque arrêt n'appartient qu'à une seule ligne. Ceci nous permet de définir pour les arrêts leur prédécesseur (arrêt précédant immédiatement l'arrêt concerné sur la même ligne) et leur successeur (arrêt suivant immédiatement l'arrêt concerné sur la même ligne). Il faut noter que les terminus de départ n'ont pas de prédécesseur et que les terminus d'arrivée n'ont pas de successeur.

Les paramètres de service des lignes que nous avons besoin de connaître sont :

. pour chaque ligne, sa fréquence et la capacité des autobus qui la desservent (capacité de places assises et capacité totale);

. pour chaque arrêt, les distributions des intervalles interautobus;

. pour chaque section de ligne (lien entre deux arrêts successifs d'une même ligne), son temps de parcours par un autobus vide ou presque vide.

4.2.4 Coefficients de coûts

Les coefficients de coûts servent à transformer les temps de parcours en coûts généralisés. Le système d'unités utilisé pour définir ces coefficients est sans importance,car c'est l'impact relatif des divers temps qui détermine l'affectation obtenue. Pour chaque application du modèle, il convient néanmoins d'évaluer les coefficients avec beaucoup de soin, sinon les résultats obtenus pourraient être assez décevants.

A cause de l'hypothèse d'homogénéité des usagers par rapport aux coûts, trois coefficients seulement suffisent dans le modèle :

. α_a, coût moyen par unité de temps passée à attendre à un arrêt;

. α_b, coût moyen par unité de temps passée à bord d'un autobus vide ou presque vide;

. α_m, coût moyen par unité de temps passée à marcher.

Si l'on désire imposer des pénalités pour les correspondances entre les lignes, on peut ajouter à la liste précédente :

. α_c, coût moyen de pénalité par correspondance.

4.2.5 Coût des déplacements à pied et des correspondances

Dans le modèle, nous faisons l'hypothèse simplificatrice que le temps moyen de parcours d'un arc de marche (déplacement à pied) est constant (et donc indépendant du nombre de personnes qui l'empruntent). Cette hypothèse nous semble justifiée par les deux faits suivants :

a) le temps requis pour effectuer un déplacement à pied n'augmente sensiblement que si la densité de piétons devient extrêmement élevée, ce qui est fort rare;

b) les arcs de marche du graphe augmenté du réseau correspondent dans la réalité à un assez grand nombre de rues et, même si plusieurs personnes empruntent un arc, cela ne veut pas dire que la densité de piétons soit forte.

Le temps de parcours des arcs de marche étant constant, leur coût l'est aussi. Il est de la forme $(\alpha_m T_a)$, où T_a est le temps de parcours.

Les arcs de correspondance sont à proprement parler des arcs de marche puisqu'ils correspondent au déplacement qu'on effectue à pied entre deux arrêts situés à proximité l'un de l'autre. Ils diffèrent cependant des autres arcs de marche à cause de la pénalité que l'on peut imposer sur les correspondances. De manière générale, on dira que leur coût est de la forme $(\alpha_c + \alpha_m T_a)$, où T_a est le temps de marche associé à l'arc.

4.2.6 Temps d'attente et capacité

Nous avons insisté à plusieurs reprises dans les chapitres précédents sur le fait que le temps d'attente aux arrêts dépendait de la congestion du

système de transport en commun; de manière plus précise, on peut dire que, à chaque arrêt, le temps d'attente moyen des passagers est une fonction croissante du taux moyen d'arrivée des passagers à cet arrêt; c'est aussi une fonction décroissante du nombre moyen de places libres à bord des autobus desservant cet arrêt. Pour la formulation mathématique du problème, il est plus simple de dire que le temps d'attente moyen à un arrêt est une fonction croissante du nombre moyen de places occupées à bord de chaque autobus (il est clair que ce nombre est égal au nombre total de places moins le nombre moyen de places libres). Le temps d'attente moyen dépend aussi des distributions de probabilité régissant les intervalles entre deux autobus successifs, le nombre de places occupées à bord des autobus et les arrivées des passagers à l'arrêt.

Dans ce chapitre, nous nous contentons d'inclure ces fonctions d'attente dans le modèle sans faire les développements mathématiques complets qui sont nécessaires pour chaque cas particulier. Dans le prochain chapitre, nous développerons certains concepts essentiels à l'obtention d'expressions mathématiques des fonctions d'attente pour certaines classes de distributions des intervalles entre les autobus, du nombre de places occupées dans les autobus et des arrivées de passagers.

Ayant déjà tenu compte de l'effet qu'a sur les passagers la capacité limitée des autobus dans la définition des fonctions d'attente, il ne nous est pas nécessaire d'ajouter des contraintes de capacité au modèle. Nous verrons d'ailleurs dans les chapitres suivants que les fonctions d'attente se trouvent à jouer le rôle de contraintes de capacité implicites dans ce modèle.

4.2.7 Coût généralisé des sections de lignes

Le coût généralisé d'une section de ligne est la somme de deux coûts de natures disparates :

a) un coût brut de parcours qui est égal au temps de parcours de la section multiplié par le coefficient α_b;

b) un coût d'inconfort qui croît rapidement quand le nombre moyen d'usagers empruntant la section s'approche ou dépasse la capacité de places assises de la ligne.

Dans la réalité, il semble que le temps de parcours d'une section de ligne donnée par un autobus dépende beaucoup plus de la congestion routière (qui n'entre pas dans le cadre du modèle) sur les rues empruntées par l'autobus que du nombre d'usagers à bord de celui-ci. Cependant, il est tout à fait raisonnable de penser que le temps de parcours va augmenter de manière légère, mais sensible, avec le nombre de passagers (temps plus long aux arrêts pour faire monter ou descendre des usagers, arrêts plus fréquents, etc.). Si on appelle t_a le temps de parcours d'une section et v_a le nombre moyen d'usagers sur celle-ci, on aurait donc une relation fonctionnelle entre t_a et v_a qui pourrait avoir l'air de ceci :

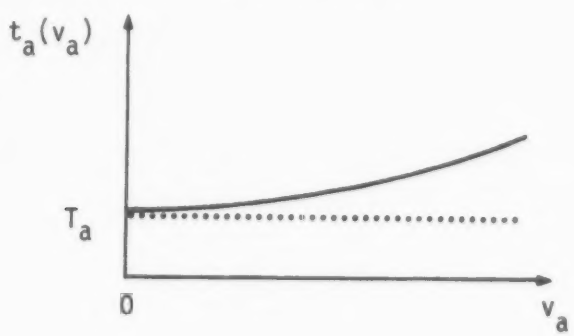


Figure 3 : Graphe typique de fonction $t_a(v_a)$.

Dans la figure 3, l'ordonnée à l'origine T_a indique le temps de parcours de la section par un autobus vide.

Pour ce qui est du coût d'inconfort, on peut supposer qu'il est à toutes fins pratiques nul quand l'autobus est assez vide, mais que, dès que l'on dépasse un certain taux d'occupation des places assises de l'autobus, il commence à croître rapidement. Implicitement, on pourrait relier ce coût à la probabilité de trouver une place assise, étant donnée la charge moyenne de la section de ligne. Si on appelle $\overline{\Psi}_a(v_a)$ ce coût et K^o_ℓ la capacité de places assises par unité de temps pour la ligne ℓ à laquelle appartient la section, on aurait quelque chose qui pourrait ressembler à ceci :

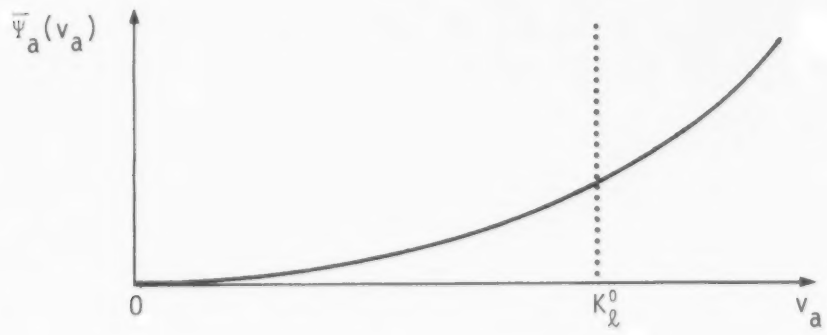


Figure 4 : Graphe typique de fonction $\overline{\Psi}_a(v_a)$.

Pour obtenir le coût total de la section de ligne (appelons-le $\tilde{\Psi}_a(v_a)$), il suffit d'ajouter le coût brut de parcours au coût d'inconfort :

$$\tilde{\Psi}_a(v_a) = \alpha_b \, t_a(v_a) + \overline{\Psi}_a(v_a)$$

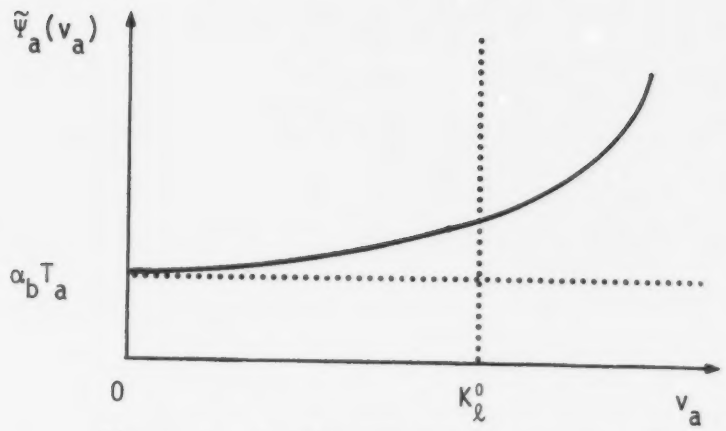


Figure 5 : Graphe typique de fonction $\tilde{\Psi}_a(v_a)$.

A ce point-ci, il faut se rendre compte que l'on ne peut pas se permettre de définir pour chacune des sections de lignes du réseau une fonction $\tilde{\Psi}_a$, car cela représenterait une tâche colossale. Nous allons donc poser un certain nombre d'hypothèses qui vont nous permettre de ramener le travail à effectuer à un niveau acceptable :

a) Des sections de lignes ayant le même temps de parcours à vide et le même taux d'occupation (rapport v_a/K_ℓ^0) ont le même temps de parcours.

b) Pour des sections de lignes ayant le même taux d'occupation, le temps de parcours est une fonction linéaire du temps de parcours à vide.

c) Des sections de lignes ayant le même temps de parcours et le même taux d'occupation ont le même coût d'inconfort.

d) Pour des sections de lignes ayant le même taux d'occupation, le coût d'inconfort est une fonction linéaire du temps de parcours à vide.

Faisons pendant un instant la supposition que toutes les lignes du réseau sont desservies par des autobus de même capacité. Dans ce cas, on peut dire,d'une certaine façon,que tous les usagers qui empruntent des lignes ayant le même taux d'occupation se trouvent dans des situations "équivalentes". C'est ce qui nous pousse à utiliser le taux d'occupation des sections comme "paramètre" dans l'énoncé de nos hypothèses. Pour ce qui concerne les temps de parcours, il nous paraît naturel de les exprimer par rapport aux temps de parcours à vide. En fait, les hypothèses a) et b) reviennent à supposer qu'il existe une fonction t (unique pour le réseau) telle que l'on puisse exprimer le temps de parcours d'une section de ligne a quelconque sous la forme :

$$t_a(v_a) = T_a \; t(v_a/K_\ell^o) \quad .$$

Arrêtons-nous maintenant à la question des coûts d'inconfort. Il semble raisonnable de penser que,pour un taux d'occupation donné, le coût d'inconfort par unité de temps soit le même pour toutes les sections de lignes. Or, c'est exactement ce que nous disent les hypothèses c) et d). Ceci pourrait aussi être exprimé sour la forme :

$$\bar{\Psi}_a(v_a) = t_a(v_a)\; \bar{\Psi}(v_a/K_\ell^o)$$

où $\bar{\Psi}$ est une fonction unique définie pour l'ensemble du réseau.

On peut maintenant réécrire $\tilde{\Psi}_a(v_a)$ en se servant des deux dernières relations. On obtient :

$$\tilde{\Psi}_a(v_a) = \alpha_b t_a(v_a) + \bar{\Psi}_a(v_a)$$

$$= \alpha_b t_a(v_a) + t_a(v_a)\bar{\Psi}(v_a/K_\ell^0)$$

$$= t_a(v_a)\,[\alpha_b + \bar{\Psi}(v_a/K_\ell^0)]$$

$$\tilde{\Psi}_a(v_a) = T_a t(v_a/K_\ell^0)\,[\alpha_b + \bar{\Psi}(v_a/K_\ell^0)]$$

Si l'on pose $\Psi(x) = t(x)\,[\alpha_b + \bar{\Psi}(x)]$, on aboutit à

$$\boxed{\tilde{\Psi}_a(v_a) = T_a\,\Psi(v_a/K_\ell^0)}$$

A partir du graphe de $\tilde{\Psi}_a(v_a)$ de la figure 5, on peut obtenir sans difficulté un graphe typique de fonction Ψ :

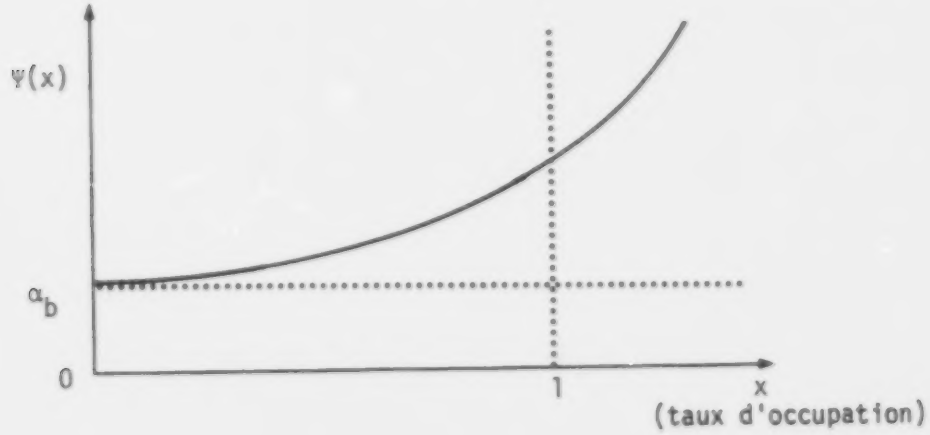


Figure 6 : Graphe typique de fonction Ψ.

Le raisonnement qui sous-tend la définition de Ψ repose sur l'hypothèse que toutes les lignes sont desservies par des autobus de même capacité. Dans le cas où on utilise dans un même réseau de transport en

commun des véhicules très différents (par exemple, si l'on a des lignes d'autobus et des lignes de métro), il faudrait sans doute définir une fonction Ψ pour chaque catégorie de véhicules. Il n'en demeure pas moins que, même pour un réseau de transport très important, on aurait au plus trois ou quatre fonctions Ψ à définir et à calibrer, ce qui est assez facile à réaliser.

4.3 Description mathématique du réseau et notation

4.3.1 Remarques générales

Le but de cette section est d'arriver à une description mathématique complète et précise du réseau de transport en commun et des autres éléments du problème d'affectation des passagers.

Le graphe augmenté du réseau (nous dirons aussi le réseau augmenté) a été défini rapidement dans la section 2.6. Précisons ici que c'est un graphe orienté [N,A] où N est l'ensemble des sommets et A celui des arcs. Les sommets représentent soit des centroïdes de zones (définis dans la section 2.6), soit des arrêts (voir la remarque au sujet des arrêts dans la sous-section 2.5). Les arcs représentent des déplacements pouvant être effectués soit à pied, soit en autobus.

Rappelons aussi que les demandes de transport en commun sont exprimées sous la forme d'une matrice origines-destinations (cf. section 2.6).

4.3.2 Définitions préliminaires

Dans le but d'éliminer toute confusion possible, nous reprenons ici les principales définitions qui sont éparpillées dans les pages qui précèdent et nous en ajoutons quelques autres qui nous seront utiles dans la suite.

Centroïde-origine : tout centroïde tel qu'il existe au moins une demande de transport ayant ce point pour origine.

Centroïde-destination : tout centroïde tel qu'il existe au moins une demande de transport ayant ce point pour destination.

Section de ligne : arc reliant deux arrêts successifs sur une même ligne et représentant un déplacement à bord d'un autobus.

Arc de marche : tout arc représentant un déplacement effectué à pied.

Arc d'accès : arc reliant un centroïde et un arrêt d'autobus.

Arc de marche directe : arc reliant deux centroïdes. Ces arcs sont inclus dans le réseau pour permettre aux usagers d'effectuer des déplacements sans avoir à utiliser le système de transport en commun proprement dit (déplacements complètement effectués à pied).

Arc de correspondance : arc reliant deux arrêts de lignes différentes, situés à proximité l'un de l'autre, afin de permettre les correspondances d'une ligne à l'autre.

Prédécesseur (d'un arrêt) : arrêt précédant immédiatement un arrêt donné sur la même ligne.

Successeur (d'un arrêt) : arrêt suivant immédiatement un arrêt donné sur la même ligne.

Terminus de départ : arrêt auquel débute une ligne; un terminus de départ n'a pas de prédécesseur.

Terminus d'arrivée : arrêt auquel se termine une ligne; un terminus d'arrivée n'a pas de successeur.

Capacité assise (d'une ligne) : nombre moyen de places assises disponibles à bord des autobus de la ligne par unité de temps.

Capacité totale (d'une ligne) : nombre moyen de places disponibles à bord des autobus de la ligne par unité de temps.

4.3.3 Notation pour le réseau augmenté

Dans l'énoncé du modèle, nous utiliserons la notation suivante :

Z, ensemble des centroïdes de zones;

R, ensemble des arrêts d'autobus;

O, ensemble des centroïdes-origines

D, ensemble des centroïdes-destinations;

S, ensemble des sections de lignes;

M, ensemble des arcs de marche;

M', ensemble des arcs de marche directe;

M", ensemble des arcs d'accès;

Q, ensemble des arcs de correspondance;

L, ensemble des lignes;

S_ℓ, ensemble des sections de ligne de la ligne ℓ;

R_ℓ, ensemble des arrêts de la ligne ℓ;

r'_ℓ, terminus de départ de la ligne ℓ;

r''_ℓ, terminus d'arrivée de la ligne ℓ;

R', ensemble des terminus de départ;

R'', ensemble des terminus d'arrivée;

K^0_ℓ, capacité assise de la ligne ℓ;

K_ℓ, capacité totale de la ligne ℓ;

$\gamma^-(r)$, prédécesseur de l'arrêt r;

$\gamma^+(r)$, successeur de l'arrêt r;

T_a, temps de parcours (à vide) de l'arc a;

g^{od}, demande de transport du centroïde-origine o vers le centroïde-destination d par unité de temps.

On pose aussi :

$$\Gamma^-(i) = \{j \in N | (j,i) \in A\} , \quad i \in N$$

$$\Gamma^+(i) = \{j \in N | (i,j) \in A\} , \quad i \in N$$

$$\overline{\Gamma}^-(r) = \Gamma^-(r) - \gamma^-(r), \quad r \in R - R'$$

$$\overline{\Gamma}^+(r) = \Gamma^+(r) - \gamma^+(r), \quad r \in R - R''$$

Note : pour les arcs de marche, T_a représente le temps effectif de parcours (fixe), tandis que, pour les sections de ligne, il représente le temps de parcours à vide; dans les deux cas, T_a est une donnée du modèle.

4.3.4 Quelques précisions

Nous allons maintenant énoncer un certain nombre de relations fondamentales du modèle, relations qui devraient permettre de clarifier les définitions et la notation que nous utilisons. Après chaque relation, nous précisons en quelques mots sa signification.

1) $N = R \cup Z$

Les sommets du graphe sont des centroïdes ou des arrêts.

2) $R \cap Z = \emptyset$

Aucun sommet n'est à la fois un centroïde et un arrêt.

3) $Z = O \cup D$

Tout centroïde est un centroïde-origine ou un centroïde-destination. Remarquons que l'on peut avoir $O \cap D \neq \emptyset$, c'est-à-dire qu'un centroïde peut être à la fois une origine et une destination.

4) $A = M \cup S$

Tout arc représente un déplacement à pied ou en autobus.

5) $M \cap S = \emptyset$

Aucun arc ne peut servir aux deux sortes de déplacements.

6) $M = M' \cup M'' \cup Q$

Les arcs de marche directe, d'accès et de correspondance sont les arcs de marche.

7) $M' \cap M'' = M' \cap Q = M'' \cap Q = \emptyset$

Chaque arc de marche n'appartient qu'à un seul type.

8) $R_{\ell_1} \cap R_{\ell_2} = \emptyset, \quad \forall\, \ell_1, \ell_2 \in L,\ \ell_1 \neq \ell_2$

Chaque arrêt n'appartient qu'à une seule ligne.

9) $S_{\ell_1} \cap S_{\ell_2} = \emptyset, \quad \forall\, \ell_1, \ell_2 \in L, \quad \ell_1 \neq \ell_2$

Chaque section de ligne n'appartient qu'à une seule ligne.

10) $R' \cap R'' = \emptyset$

Un arrêt ne peut être à la fois terminus de départ et terminus d'arrivée.

4.3.5 Eclatement des arrêts

Nous avons déjà dit (sous-section 4.2.6) que le temps d'attente moyen à un arrêt est une fonction du nombre moyen d'usagers se présentant à l'arrêt et du nombre moyen de places occupées dans les autobus de la ligne desservant l'arrêt. Il serait assez difficile d'identifier simplement ces quantités avec la structure que nous avons définie. En effet, cette structure ne nous permet pas de voir exactement ce qui se passe à un arrêt. Pour résoudre ce problème, nous allons éclater chaque arrêt en trois sommets. Les figures 7 et 8 permettent de voir les modifications apportées par l'éclatement à la représentation d'un arrêt r donné.

Le réseau obtenu après éclatement des arrêts est appelé réseau éclaté, on le note [N*,A*].

Avec la représentation éclatée, il suffit de définir de manière adéquate les unités dans lesquelles sont exprimées les variables de flots du modèle, pour obtenir les correspondances suivantes :

a) Le nombre moyen de places occupées à bord des autobus passant à l'arrêt r (par unité de temps) est donné par le flot total d'usagers sur l'arc auxiliaire de l'arrêt r (ce flot est noté $\bar{v}_r$).

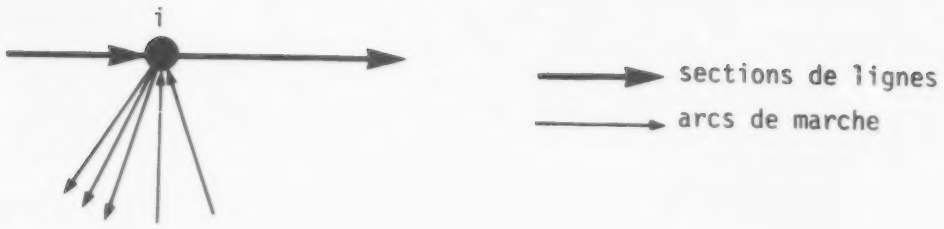


Figure 7 : L'arrêt simple r en représentation non éclatée.

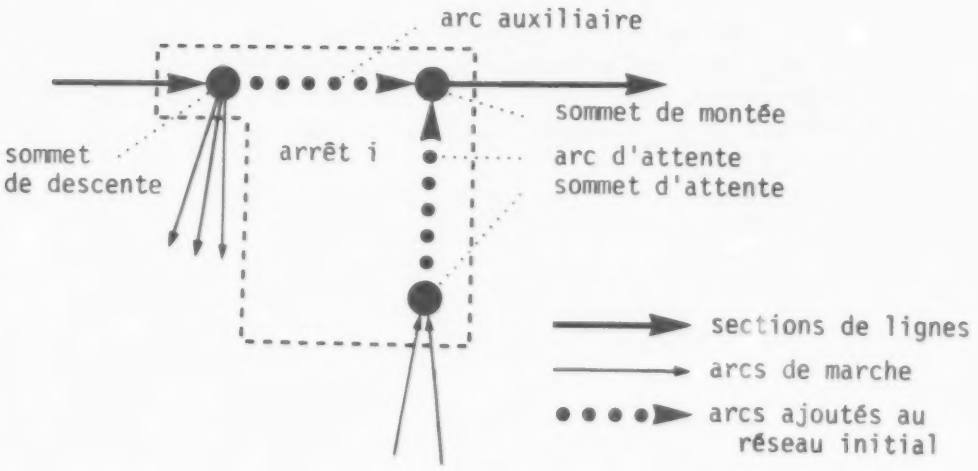


Figure 8 : L'arrêt simple r en représentation éclatée.

b) Le nombre moyen d'usagers se présentant à l'arrêt r (par unité de temps) est donné par le flot total d'usagers sur l'arc d'attente de l'arrêt r (ce flot est noté $\tilde{v}_r$).

Avant de poursuivre, il faut remarquer que $\bar{v}_r$ n'est pas le nombre moyen de places occupées par autobus passant à l'arrêt r; $\bar{v}_r$ est plutôt

égal à ce dernier nombre multiplié par la fréquence de la ligne desservant l'arrêt r.

Pour ce qui est des coûts de parcours, on peut noter que les arcs auxiliaires ne représentent aucun déplacement; leur temps de parcours (leur coût aussi) est donc nul. Les arcs d'attente pour leur part représentent les mouvements dans les files d'attente aux arrêts; il est donc logique de leur associer comme temps de parcours les temps d'attente aux arrêts. Pour chaque arrêt r, le temps d'attente est noté $\Phi_r(\bar{v}_r, \tilde{v}_r)$. La fonction Φ_r est définie à partir de la distribution des intervalles interautobus à l'arrêt r, de la distribution choisie pour décrire le processus d'arrivée des usagers à l'arrêt et de la capacité totale des autobus de la ligne desservant l'arrêt.

Avant de terminer cette sous-section, il faut remarquer que les terminus de départ ne sont éclatés qu'en deux sommets (sommets d'attente et d'embarquement) et que les terminus d'arrivée ne sont représentés que par un sommet de descente, car, dans chacun des cas, les autres sommets ne nous seraient d'aucune utilité.

4.3.6 Notation pour le réseau éclaté

Il n'est pas nécessaire de redéfinir complètement une nouvelle notation pour le réseau éclaté. Il nous semble plus avantageux de conserver au maximum la notation du réseau augmenté, car elle fait apparaître clairement les arrêts comme des touts structurés. Pour éviter toute confusion, il suffit de se rappeler que, dans le réseau éclaté, tous les arcs de marche qui "partent" d'un arrêt et la section de ligne qui "vient"

du prédécesseur de l'arrêt sont incidents au sommet de descente, que tous les arcs de marche qui "arrivent" à l'arrêt sont incidents au sommet d'attente, et que la section de ligne qui "va" vers le successeur de l'arrêt est incidente au sommet d'embarquement.

Pour les cas où il nous sera nécessaire de faire référence directement aux sommets d'attente, de descente et d'embarquement, nous définissons les notations additionnelles suivantes :

$u(r)$, sommet d'attente de l'arrêt $r, r \in R\text{-}R''$;

r^-, sommet de descente de l'arrêt $r, r \in R\text{-}R'$;

r^+, sommet d'embarquement de l'arrêt $r, r \in R\text{-}R''$;

U, ensemble des sommets d'attente;

R^-, ensemble des sommets de descente;

R^+, ensemble des sommets d'embarquement;

$\tilde{\Gamma}^+(i) = \{j \in N^* | (i,j) \in A^*\}$;

$\tilde{\Gamma}^-(i) = \{j \in N^* | (j,i) \in A^*\}$.

Avec cette notation, on peut identifier l'ensemble des arcs auxiliaires à l'intersection de A^* et du produit cartésien $R^- \times R^+$ ainsi que l'ensemble des arcs d'attente à l'intersection de A^* et du produit cartésien $U \times R^+$.

4.3.7 Variables de flot

Nous avons vu dans la section 2.9 qu'il était nécessaire de différencier sur chaque arc les flots d'usagers soit par paire O-D, soit par destination, soit par origine. Même si,dans la majorité des modèles d'affectation des passagers, on utilise une différenciation par origine, dans ce modèle nous distinguerons les usagers par destination, car cela nous sera absolument nécessaire quand nous voudrons ajouter des arrêts multiples au modèle dans le chapitre sixième. Nous utiliserons donc les variables de flot suivantes :

v_{ij}^{d}, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc (i,j) (où (i,j) est un arc de A, c'est-à-dire que i ou j peut être un arrêt);

$\bar{v}_{r}^{d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc auxiliaire de l'arrêt r;

$\tilde{v}_{r}^{d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc d'attente de l'arrêt r.

A partir de ces variables, il est facile d'obtenir les flots totaux sur les arcs. Si on dénote respectivement par v_{ij}, $\bar{v}_{r}$ et $\tilde{v}_{r}$ les flots totaux sur les arcs de A, les arcs auxiliaires et les arcs d'attente, on peut écrire les relations suivantes :

$$v_{ij} = \sum_{d \in D} v_{ij}^{d} , \qquad \forall\ (i,j) \in A$$

$$\bar{v}_{r} = \sum_{d \in D} \bar{v}_{r}^{d} , \qquad \forall\ r \in R-R'$$

$$\tilde{v}_r = \sum_{d \in D} \tilde{v}_r^d , \qquad \forall\, r \in R-R''$$

Les quantités v_{ij}, $\bar{v}_r$ et $\tilde{v}_r$ ne représentent pas à proprement parler des variables du modèle, on ne fait que les déduire des valeurs des variables.

4.3.8 Fonctions de coût sur les arcs

Pour simplifier l'énoncé de la condition d'équilibre (section 4.5), nous allons définir une notation commune pour les coûts de parcours. Comme les coûts de plusieurs arcs du réseau dépendent des flots totaux sur un ou plusieurs arcs, nous allons souligner cette dépendance dans la notation choisie. Soit $\underline{v}$ un vecteur de flots totaux sur le réseau éclaté ($\underline{v} = (v_a)_{a \in A^*}$), on appelle $c_a(\underline{v})$ le coût induit par $\underline{v}$ pour l'arc a ($a \in A^*$) et on pose $\underline{c}(\underline{v}) = (c_a(\underline{v}))_{a \in A^*}$.

Avec cette notation, on peut réécrire le coût des arcs du réseau à partir des résultats des sous-sections 4.2.5, 4.2.7 et 4.3.5. On obtient :

. pour les arcs d'accès et de marche directe

$$c_a(\underline{v}) = \alpha_m T_a, \quad a \in M' \cup M''$$

. pour les arcs de correspondance

$$c_a(\underline{v}) = \alpha_c + \alpha_m T_a, \quad a \in Q$$

. pour les sections de lignes

$$c_a(\underline{v}) = T_a \, \Psi \, (v_a / K_\ell^0), \quad a \in S_\ell, \quad \ell \in L$$

. pour les arcs auxiliaires

$$c_a(\underline{v}) = 0, \qquad a \in R^- \times R^+$$

. pour les arcs d'attente des arrêts qui ne sont pas des terminus

$$c_a(\underline{v}) = \alpha_a \, \Phi_r \, (\bar{v}_r, \tilde{v}_r), \quad a = (u(r), r^+), \quad r \in R - (R' \cup R'')$$

. pour les arcs d'attente des terminus de départ

$$c_a(\underline{v}) = \alpha_a \, \Phi_r \, (0, \tilde{v}_r), \quad a = (u(r), r^+), \quad r \in R'$$

4.4 Affectations réalisables

Nous allons maintenant caractériser les affectations réalisables des passagers, c'est-à-dire celles qui sont possibles en pratique. Nous utiliserons une formulation sommets-arcs qui permet de bien mettre en évidence les divers flots d'usagers sur les arcs du réseau. Une telle formulation a le défaut de ne pas nous donner directement les chemins (itinéraires) utilisés par les usagers, mais nous verrons à la section 4.5 que cela ne nous gênera pas dans l'énoncé de la condition d'équilibre du modèle.

Pour qu'une affectation soit réalisable, il suffit que les demandes de transport soient satisfaites et que les équations de conservation de flot soient aussi satisfaites à tous les sommets du réseau. En utilisant la notation définie pour le réseau éclaté, on obtient le système d'équations et d'inéquations suivant :

$$(1') \qquad \sum_{j \in \tilde{\Gamma}^+(d)} v^d_{dj} - \sum_{j \in \tilde{\Gamma}^-(d)} v^d_{jd} = - \sum_{o \in O} g^{od}, \qquad d \in D$$

$$(2') \qquad \sum_{j \in \tilde{\Gamma}^+(o)} v^d_{oj} - \sum_{j \in \tilde{\Gamma}^-(o)} v^d_{jo} = g^{od}, \qquad d \in D, \; o \in O - \{d\}$$

$$(3') \quad \sum_{j \in \tilde{\Gamma}^+(i)} v_{ij}^d - \sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d = 0, \quad d \in D,\ i \in N^* - (O \cup \{d\})$$

$$(4') \quad v_{ij}^d \geq 0, \quad d \in D,\ (i,j) \in A^*$$

Les équations (1') et (2') nous assurent que les demandes de transport sont satisfaites respectivement aux destinations et aux origines. Dans l'équation (2'), on pose $o \in O - \{d\}$, car d peut appartenir à O.

Les équations (3') sont les équations de conservation de flot pour les autres sommets du réseau.

Les inéquations (4') indiquent tout simplement que les flots d'usagers ne peuvent être négatifs.

Ce premier système d'inéquations est compact, mais il ne nous permet pas de voir très bien ce qui se passe pour chaque type de sommets. Nous allons donc réécrire les conditions que doit vérifier une affectation pour être réalisable en utilisant au maximum la notation définie par le réseau augmenté et en distinguant les types de sommets. On obtient alors :

$$(1) \quad \sum_{j \in \Gamma^+(d)} v_{dj}^d - \sum_{j \in \Gamma^-(d)} v_{jd}^d = - \sum_{o \in O} g^{od}, \qquad d \in D$$

$$(2) \quad \sum_{j \in \Gamma^+(o)} v_{oj}^d - \sum_{j \in \Gamma^-(o)} v_{jo}^d = g^{od}, \qquad d \in D,\ o \in O - \{d\}$$

$$(3) \quad \sum_{j \in \Gamma^+(z)} v_{zj}^d - \sum_{j \in \Gamma^-(z)} v_{jz}^d = 0, \qquad d \in D,\ z \in Z - (O \cup \{d\})$$

$$(4) \quad \sum_{j \in \Gamma^+(r)} v^d_{rj} + \bar{v}^d_r - v^d_{\gamma^-(r),r} = 0, \qquad d \in D,\ r \in R - (R' \cup R'')$$

$$(5) \quad \tilde{v}^d_r - \sum_{j \in \Gamma^-(r)} v^d_{jr} = 0, \qquad d \in D,\ r \in R - (R' \cup R'')$$

$$(6) \quad v^d_{r,\gamma^+(r)} - \bar{v}^d_r - \tilde{v}^d_r = 0, \qquad d \in D,\ r \in R - (R' \cup R'')$$

$$(7) \quad \tilde{v}^d_r - \sum_{j \in \Gamma^-(r)} v^d_{jr} = 0, \qquad d \in D,\ r \in R'$$

$$(8) \quad v^d_{r,\gamma^+(r)} - \tilde{v}^d_r = 0, \qquad d \in D,\ r \in R'$$

$$(9) \quad \sum_{j \in \Gamma^+(r)} v^d_{rj} - v^d_{\gamma^-(r),r} = 0, \qquad d \in D,\ r \in R''$$

$$(10) \quad v^d_{ij} \geq 0, \qquad d \in D,\ (i,j) \in A$$

$$(11) \quad \bar{v}^d_r \geq 0, \qquad d \in D,\ r \in R - (R' \cup R'')$$

$$(12) \quad \tilde{v}^d_r \geq 0, \qquad d \in D,\ r \in R - R''$$

Les équations (1) et (2) correspondent respectivement aux équations (1') et (2') du système précédent.

Les contraintes (3) sont les équations de conservation de flot aux centroïdes qui ne sont ni des origines, ni la destination considérée.

Les contraintes (4), (5) et (6) sont, dans l'ordre, les équations de conservation de flot aux sommets de descente, d'attente et d'embarquement des arrêts qui ne sont pas des terminus.

Les équations (7) et (8) sont les équations de conservation de flot aux sommets d'attente et d'embarquement des terminus de départ.

Les contraintes (9) sont les équations de conservation de flot aux sommets de descente des terminus d'arrivée.

Les inéquations (10), (11) et (12) correspondent aux inéquations (4') de la formulation compacte.

Pour bien comprendre ce que représentent les contraintes (4) et (5), il faut remarquer que, pour tout arrêt r qui n'est pas un terminus, on a :

$j \in \overline{\Gamma}^+(r)$ <=> (r,j) est un arc de marche incident au sommet de descente de r;

$j \in \overline{\Gamma}^-(r)$ <=> (j,r) est un arc de marche incident au sommet d'attente de r.

Pour les terminus d'arrivée, on a :

$j \in \Gamma^+(r)$ <=> (r,j) est un arc de marche incident au sommet de descente de r, car r n'a pas de successeur.

Similairement, pour les terminus de départ, on a :

$j \in \Gamma^-(r)$ <=> (j,r) est un arc de marche incident au sommet d'attente de r, car r n'a pas de précédesseur.

4.5 Affectations d'équilibre

Le but de cette section est de déterminer quelles conditions doit remplir une affectation réalisable pour satisfaire l'hypothèse d'optimisation égoïste.

4.5.1 Interprétations diverses de l'optimisation égoïste

Dans son énoncé le plus simple, l'hypothèse d'optimisation égoïste est quelque peu imprécise : on suppose bien sûr que chaque usager essaie de minimiser son coût de transport personnel, mais on ne spécifie pas les informations sur lesquelles sont fondés les choix d'itinéraires (ou de stratégies).

Smith (1979) a proposé un schéma dynamique dans lequel l'hypothèse de comportement peut être interprétée de diverses façons. Ce schéma repose sur la supposition que les usagers doivent accomplir jour après jour les mêmes déplacements. A chaque jour, les usagers peuvent utiliser le même itinéraire que la veille ou un itinéraire moins coûteux. On atteint un point d'équilibre quand plus aucun usager n'a la possibilité de changer d'itinéraire pour un qui serait moins coûteux. Il y a une imprécision qui subsiste ici quand nous parlons d'"itinéraire moins coûteux" : chaque affectation induisant un vecteur de coûts pour les arcs du réseau, le coût d'un itinéraire varie avec les affectations. Smith suggère d'utiliser les coûts induits par l'affectation d'hier pour déterminer les changements d'itinéraires possibles aujourd'hui, car ces coûts constituent l'information disponible la plus récente. Il serait possible de baser les changements d'itinéraires sur les coûts qui en résulteront aujourd'hui,

mais cela impliquerait que les usagers puissent prévoir comment leurs changements d'itinéraires influeront sur les coûts. Pour diverses raisons pratiques, il nous paraît plus simple de retenir la suggestion de Smith et de travailler avec les coûts de l'affectation d'hier.

Notons, en passant, que Smith travaille pour sa part avec une interprétation "globalisée" de l'hypothèse de comportement, c'est-à-dire qu'au lieu de ne permettre que les changements d'itinéraires qui améliorent le sort de chacun des usagers impliqués (par rapport aux coûts d'hier, toujours), il considère comme acceptables toutes les affectations pour lesquelles le coût total de transport (pour l'ensemble des usagers) est plus faible que celui de l'affectation d'hier. C'est cette interprétation "globalisée" qui permet de reformuler le problème de recherche d'une affectation d'équilibre sous la forme d'un problème d'inéquation variationnelle.

4.5.2 Condition d'équilibre

Supposons que l'on ait en main une affectation réalisable des usagers et que l'on veuille déterminer si c'est une affectation d'équilibre. Pour ce faire, il suffit de considérer cette affectation d'hier et de regarder si certains des usagers auraient avantage à changer d'itinéraire. Cela peut se produire si et seulement si certains des itinéraires utilisés ne correspondent pas à un chemin le plus court (en termes des coûts d'hier) entre l'origine et la destination des usagers impliqués. La façon la plus simple de vérifier cette condition consiste à calculer pour chaque destination le sous-graphe des chemins les plus courts de tous les sommets du

réseau (éclaté) vers cette destination, et à regarder s'il existe une destination pour laquelle il y a un flot non-nul d'usagers (se dirigeant vers elle) sur un arc n'appartenant pas au sous-graphe des chemins les plus courts. Dans un tel cas, on peut affirmer que les itinéraires des usagers empruntant cet arc ne sont pas des chemins les plus courts vers la destination en question et que, donc, l'affectation en main n'est pas une affectation d'équilibre. Si, d'autre part, pour toute destination, les flots d'usagers (se dirigeant vers elle) sont nuls sur tous les arcs n'appartenant pas au sous-graphe des chemins les plus courts (vers cette destination), on peut en conclure qu'aucun usager ne peut réduire son coût de déplacement (par rapport au vecteur de coûts induit par l'affectation en main), que, pour chaque paire origine-destination, le coût des itinéraires utilisés est égal (c'est la "longueur" des chemins les plus courts de l'origine vers la destination) et qu'il est inférieur à celui des itinéraires non-utilisés.

Caractérisons maintenant cette condition en termes mathématiques. Soit $\{v_a^{d*}\}_{a\in A^*}^{d\in D}$ l'affectation des passagers en main, $\underline{v}^*$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v}^*)$ le vecteur des coûts induits par ces flots totaux et $(\Pi_i^d(\underline{v}^*))_{i\in N^*}^{d\in D}$ le vecteur des "longueurs" des chemins les plus courts de tous les sommets vers toutes les destinations pour le vecteur de flots v^*, on a alors les relations suivantes :

1) $\forall\ d\in D, \quad \Pi_d^d(\underline{v}^*) = 0$

2) $\forall\ d\in D, \ \forall\ i\in N^* - \{d\}, \ \Pi_i^d(\underline{v}^*) = \min_{j\in\tilde{\Gamma}^+(i)} \{c_{ij}(\underline{v}^*) + \Pi_j^d(\underline{v}^*)\}$

3) $\forall\ d\in D, \ \forall\ (i,j)\in A^*,$

$< c_{ij}(\underline{v}^*) = \Pi_i^d(\underline{v}^*) - \Pi_j^d(v^*) >$ <=> (i,j) appartient au sous-graphe des chemins les plus courts vers la destination d

Il faut noter que ces propriétés ne sont vraies (en toute rigueur) que si le réseau éclaté est fortement connexe par rapport aux destinations, c'est-à-dire que s'il existe de tout sommet i du réseau un chemin vers toute destination d. C'est une hypothèse peu restrictive qui devrait toujours être vérifiée en pratique. Dans la suite, nous supposerons implicitement qu'elle l'est.

A partir des relations précédentes, nous pouvons maintenant écrire la condition d'équilibre du modèle.

Théorème 4.1 : Condition d'équilibre

Soit $\{v_a^{d}*\}_{a\in A^*}^{d\in D}$ une affectation réalisable des usagers, $\underline{v}^*$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v}^*)$ le vecteur de coûts induit par ces flots totaux et $(\Pi_i^d(\underline{v}^*))_{i\in N^*}^{d\in D}$ le vecteur des "longueurs" des chemins les plus courts pour $\underline{v}^*$, $\{v_a^{d}*\}$ est une affectation d'équilibre si et seulement si

$$\forall\ (i,j)\in A^*,\ \forall\ d\in D,\quad v_{ij}^{d}*(\Pi_i^d(\underline{v}^*)-\Pi_j^d(\underline{v}^*)-c_{ij}(\underline{v}^*)) = 0 \quad .$$

Démonstration

L'affectation $\{v_a^{d}*\}$ étant réalisable, on a toujours pour un arc (i,j) et une destination d donnés $v_{ij}^{d}* \geq 0$ et la condition du théorème est alors équivalente à

$$< v_{ij}^d * > 0 \Rightarrow \pi_i^d(\underline{v}^*) - \pi_j^d(\underline{v}^*) - c_{ij}(\underline{v}^*) = 0 >$$

$$\Leftrightarrow \quad < v_{ij}^d * > 0 \Rightarrow c_{ij}(\underline{v}^*) = \pi_i^d(\underline{v}^*) - \pi_j^d(\underline{v}^*) >$$

$$\Leftrightarrow \quad < v_{ij}^d * > 0 \Rightarrow (i,j) \text{ appartient au sous-graphe des chemins les plus courts vers d} >$$

C'est donc dire que cette condition est équivalente à

$$(i,j) \text{ n'appartient pas au sous-graphe des chemins les plus courts vers d} \Rightarrow v_{ij}^d * = 0$$

et $\{v_a^d *\}$ est bien une affectation d'équilibre tel qu'expliqué auparavant. □

4.5.3 Reformulation du principe de Wardrop

Dans son énoncé original, le premier principe de Wardrop correspond à une formulation arcs-chemins du problème d'affectation des automobilistes, tandis que la formulation de Smith, sous forme d'inéquation variationnelle, repose sur l'interprétation "globalisée" de l'hypothèse de comportement. La condition d'équilibre de notre modèle, qui correspond à une formulation sommets-arcs du problème d'affectation, constitue donc une manière nouvelle d'énoncer le principe de Wardrop. On peut l'écrire sous la forme suivante :

> Les itinéraires utilisés par les usagers sont formés uniquement d'arcs appartenant au sous-graphe des chemins les plus courts vers leur destination.

4.5.4 Mesure de déséquilibre

Etant donnée une affectation réalisable, il peut être intéressant d'avoir une mesure quelconque nous permettant de voir si c'est une

affectation d'équilibre et,si ce n'en est pas une, d'avoir une certaine idée du degré de "déséquilibre" de celle-ci. Pour ce faire, nous allons utiliser la propriété suivante :

$$\forall\, \underline{v},\ \forall (i,j)\in A^*,\ \forall d\in D,\ c_{ij}(\underline{v}) - (\pi_i^d(\underline{v}) - \pi_j^d(\underline{v})) \geq 0 \ .$$

Définition 4.1 : Soit $\{v_a^d\}_{a\in A^*}^{d\in D}$ une affectation réalisable quelconque, $\underline{v}$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v})$ le vecteur de coûts induit par ces flots totaux et $(\pi_i^d(\underline{v}))_{i\in N^*}^{d\in D}$ le vecteur des "longueurs" des chemins les plus courts pour $\underline{v}$, on appelle mesure de déséquilibre de $\{v_a^d\}$ la quantité $\tau\{v_a^d\}$ suivante :

$$\tau\{v_a^d\} = \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d (c_{ij}(\underline{v}) - (\pi_i^d(\underline{v}) - \pi_j^d(\underline{v})) \ .$$

Propriété 4.1 : La mesure de déséquilibre de toute affectation réalisable est non-négative.

Propriété 4.2 : La mesure de déséquilibre d'une affectation réalisable est nulle si et seulement si c'est une affectation d'équilibre.

La propriété 4.1 découle de la propriété citée au début de la sous-section et de la non-négativité des flots.

La propriété 4.2 découle du fait que la mesure de déséquilibre étant une somme de termes non-négatifs, elle ne peut être nulle que si tous ces termes sont nuls, et alors les conditions du théorème 4.1 sont satisfaites.

Théorème 4.2 : Soit $\{v_a^d\}_{a\in A^*}^{d\in D}$ une affectation réalisable, $\underline{v}$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v})$ le vecteur de coûts induit par $\underline{v}$ et $(\pi_i^d(\underline{v}))_{i\in N^*}^{d\in D}$ le vecteur des "longueurs" des chemins les plus courts pour $\underline{v}$, alors on a

$$\tau\{v_a^d\} = \sum_{(i,j)\in A^*} c_{ij}(\underline{v})v_{ij} - \sum_{d\in D}\sum_{o\in O} \pi_o^d(\underline{v})g^{od} \quad .$$

Démonstration

On a

$$\tau\{v_a^d\} = \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \,(c_{ij}(\underline{v})-(\pi_i^d(\underline{v}) - \pi_j^d(\underline{v}))$$

$$\Rightarrow \quad \tau\{v_a^d\} = \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d\, c_{ij}(\underline{v}) - \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d\, \pi_i^d(\underline{v})$$

$$+ \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d\, \pi_j^d(\underline{v}) \qquad (0)$$

Examinons maintenant chacune des doubles sommations séparément. On a

$$\sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d c_{ij}(\underline{v}) = \sum_{(i,j)\in A^*} c_{ij}(\underline{v}) \sum_{d\in D} v_{ij}^d = \sum_{(i,j)\in A^*} c_{ij}(\underline{v})v_{ij} \qquad (1)$$

par la définition de $\underline{v}$.

D'autre part,

$$\sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \pi_i^d(\underline{v}) = \sum_{d\in D}\sum_{i\in N^*} \pi_i^d(\underline{v}) \sum_{j\in\tilde{\Gamma}^+(i)} v_{ij}^d \qquad (2)$$

et

$$\sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \pi_j^d(\underline{v}) = \sum_{d\in D}\sum_{j\in N^*} \pi_j^d(\underline{v}) \sum_{i\in\tilde{\Gamma}^-(j)} v_{ij}^d \qquad (3)$$

par les identités

$$A^* = \bigcup_{i \in N^*} \{(i,j) | j \in \tilde{\Gamma}^+(i)\} = \bigcup_{j \in N^*} \{(i,j) | i \in \tilde{\Gamma}^-(j)\}$$

En interchangeant le rôle des indices dans le membre droit de (3), on obtient

$$\sum_{d \in D} \sum_{(i,j) \in A^*} v_{ij}^d \pi_j^d(\underline{v}) = \sum_{d \in D} \sum_{i \in N^*} \pi_i^d(\underline{v}) \sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d \qquad (4)$$

En combinant (2) et (4), on obtient :

$$\sum_{d \in D} \sum_{(i,j) \in A^*} v_{ij}^d \pi_j^d(\underline{v}) - \sum_{d \in D} \sum_{(i,j) \in A^*} v_{ij}^d \pi_i^d(\underline{v}) =$$

$$\sum_{d \in D} \sum_{i \in N^*} \pi_i^d(\underline{v}) \sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{d \in D} \sum_{i \in N^*} \pi_i^d(\underline{v}) \sum_{j \in \tilde{\Gamma}^+(i)} v_{ij}^d =$$

$$\sum_{d \in D} \sum_{i \in N^*} \pi_i^d(\underline{v}) \left[\sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j \in \tilde{\Gamma}^+(i)} v_{ij}^d \right] \qquad (5)$$

Pour $i \in N^* - (O \cup \{d\})$, on a

$$\sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j \in \tilde{\Gamma}^+(i)} v_{ij}^d = 0 \quad , \quad \forall\, d$$

(contrainte (3') du système définissant les affectations réalisables).

Pour $i \in O - \{d\}$, on a

$$\sum_{j \in \tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j \in \tilde{\Gamma}^+(i)} v_{ij}^d = -g^{id} \quad , \quad \forall\, d$$

(contrainte (2') du même système).

D'autre part, pour i = d, on a

$$\pi_d^d(\underline{v}) = 0 \quad , \qquad \forall\, d$$

(Définition de $(\pi_i^d(\underline{v}))$) .

Si $d \in O$, on devrait avoir $g^{dd} = 0$, mais de toute façon rien ne nous empêche de poser

$$\pi_d^d(\underline{v}) \left[\sum_{j \in \tilde{\Gamma}^-(d)} v_{jd}^d - \sum_{j \in \tilde{\Gamma}^+(d)} v_{dj}^d \right] = -\pi_d^d(\underline{v}) g^{dd}$$

En définitive, l'équation (5) est donc équivalente à

$$\sum_{d \in D} \sum_{(i,j) \in A^*} v_{ij}^d \pi_j^d(\underline{v}) - \sum_{d \in D} \sum_{(i,j) \in A^*} v_{ij}^d \pi_i^d(\underline{v}) =$$

$$- \sum_{d \in D} \sum_{o \in O} \pi_o^d(\underline{v}) g^{od} \qquad (6)$$

En substituant (1) et (6) dans (0), on obtient

$$\tau\{v_a^d\} = \sum_{(i,j) \in A^*} c_{ij}(\underline{v}) v_{ij} - \sum_{d \in D} \sum_{o \in O} \pi_o^d(\underline{v}) g^{od} \, . \qquad \square$$

Corollaire 4.1 : Deux affectations réalisables engendrant le même vecteur de flots totaux ont la même mesure de déséquilibre.

Le théorème 4.2 fait aussi apparaître que la notion de mesure de déséquilibre est très proche de celle de fonction d'écart définie par Hearn (1982) (voir aussi Hearn, Lawphongpanich et Nguyen, 1983) ; en général, pour un problème donné, les deux concordent.

4.5.5 Equilibre et flots totaux

Suite aux résultats du théorème 4.2, il devient intéressant de caractériser la notion d'équilibre par rapport aux vecteurs de flots totaux.

Définition 4.2 : Un vecteur de flots totaux $\underline{v}$ est dit vecteur de flots totaux réalisables pour un problème donné s'il existe au moins une affectation réalisable pour ce problème qui engendre $\underline{v}$.

Définition 4.3 : Soit $\underline{v}$ un vecteur de flots totaux réalisable, $\underline{c}(\underline{v})$ le vecteur de coûts induit par $\underline{v}$ et $(\Pi_i^d(\underline{v}))_{i\in N^*}^{d\in D}$ le vecteur des "longueurs" des chemins les plus courts pour $\underline{v}$, on appelle mesure de déséquilibre de $\underline{v}$, la quantité $\tau(\underline{v})$ suivante :

$$\tau(\underline{v}) = \sum_{(i,j)\in A^*} c_{ij}(\underline{v})v_{ij} - \sum_{d\in D}\sum_{o\in O} \Pi_o^d(\underline{v})g^{od}$$

Définition 4.4 : Un vecteur de flots totaux réalisable $\underline{v}$ est dit vecteur de flots totaux d'équilibre si toutes les affectations réalisables qui l'engendrent sont des affectations d'équilibre.

Théorème 4.3 : Un vecteur de flots totaux réalisable $\underline{v}$ est un vecteur de flots totaux d'équilibre si et seulement si

$$\tau(\underline{v}) = 0 \quad .$$

Démonstration

Par la définition 4.3, le théorème 4.2 et le corollaire 4.1, il est clair que

$$\tau(\underline{v}) = 0 \iff \tau\{v_a^d\} = 0 \quad \text{pour toute affectation réalisable } \{v_a^d\} \text{ engendrant } \underline{v}.$$

D'autre part, par la propriété 4.2, on sait que pour toute affectation réalisable $\{v_a^d\}$,

$$\tau\{v_a^d\} = 0 \iff \{v_a^d\} \text{ est une affectation d'équilibre.}$$

Et donc,

$$\tau(\underline{v}) = 0 \iff \text{Toute affectation réalisable engendrant } \underline{v} \text{ est une affectation d'équilibre.}$$

□

Le théorème 4.3 nous montre que, si on a une manière efficace de caractériser les vecteurs de flots totaux réalisables, on peut se contenter de travailler avec les flots totaux pour chercher un point d'équilibre.

4.5.6 Equilibre basé sur les "coûts d'aujourd'hui"

Dans notre mémoire de maîtrise (Gendreau, 1977), nous avons utilisé l'interprétation en termes de "coûts d'aujourd'hui" de l'hypothèse d'optimisation égoïste. Il est beaucoup plus difficile de travailler avec cette interprétation, car on ne peut pas conclure *a priori* sur l'impact qu'auront sur les coûts certains changements d'itinéraires. Le plus gros du travail consiste donc alors à trouver un ensemble de propriétés que doit posséder un problème pour que l'on puisse énoncer une condition d'équilibre simple. Dans le cas où cela est possible, la condition d'équilibre que l'on obtient est, à toutes fins pratiques, équivalente à celle du théorème 4.1. En définitive, il ne nous paraît pas très fructueux

d'utiliser l'interprétation en termes des "coûts d'aujourd'hui" : cela complique le travail, sans amener de résultats vraiment différents.

4.6 Existence et unicité d'une solution d'équilibre

Il nous paraît prématuré de s'interroger sur l'existence et l'unicité d'une solution d'équilibre pour le modèle avant d'avoir précisé les fonctions de coûts pour les arcs d'attente; nous traiterons donc cette question au chapitre neuvième. Nous aimerions cependant souligner deux aspects de la question dans ce chapitre-ci. Le premier est l'importance d'avoir des coûts non-constants sur les sections de lignes si l'on veut avoir l'unicité de la solution d'équilibre (voir section 4.7). Le second découle des résultats des sous-sections 4.5.4 et 4.5.5 : puisque toutes les affectations réalisables qui engendrent le même vecteur de flots totaux ont la même mesure de déséquilibre, il est à peu près certain qu'à partir d'une affectation d'équilibre donnée, on puisse, par certaines manipulations sur les flots, en trouver plusieurs autres; c'est donc dire que la question d'unicité d'une solution d'équilibre devrait être formulée par rapport aux vecteurs de flots totaux et non par rapport aux affectations proprement dites.

4.7 Importance des coûts non-constants pour les sections de lignes

Dans le but de simplifier légèrement le modèle, on peut être tenté d'utiliser des coûts constants pour les arcs représentant des sections de lignes. Malheureusement, cela peut facilement avoir pour effet de produire des "équilibres" multiples, à un tel point d'ailleurs que l'idée même d'équilibre devient un concept sans substance. C'est ce que nous allons voir dans l'exemple qui suit.

Exemple 4.1

Considérons un problème d'affectation des passagers pour lequel le réseau augmenté est le suivant :

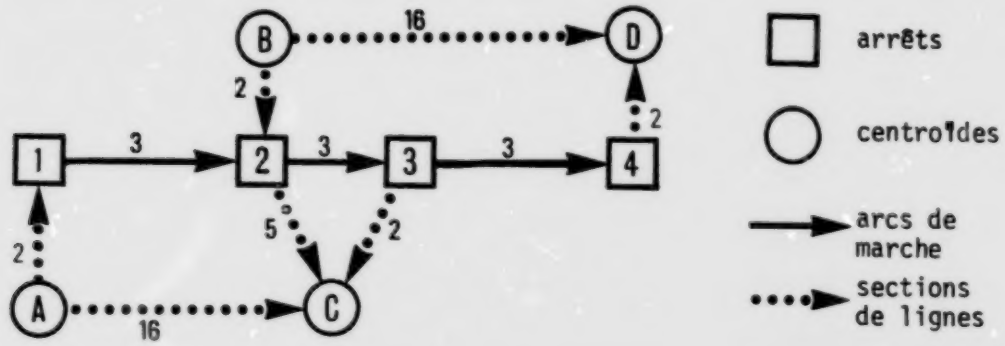


Figure 9 : Réseau de l'exemple 4.1

Dans la figure 9, les nombres placés à côté des arcs indiquent les coûts qui leur sont associés.

Pour les arcs d'attente des arrêts 1 et 2, on suppose que le coût de ces arcs est donné par :

$$c_{u(i),i^+}(\underline{v}) = 4\tilde{v}_i^2$$

et $$c_{u(2),2^+}(\underline{v}) = (\tilde{v}_2 + 2\tilde{v}_2)^2 .$$

Les seules demandes de transport non-nulles sont :

$$g^{AC} = 1 \quad \text{et} \quad g^{BD} = 1.$$

Considérons maintenant les affectations réalisables $\{v_a^{d'}\}$ et $\{v_a^{d''}\}$ données par le tableau suivant :

Arc A	$v_a^{C'}$	$v_a^{D'}$	$v_a^{C''}$	$v_a^{D''}$
(A,C)	0	0	0	0
(A,1)	1	0	1	0
(B,D)	0	0	0	$(3-\sqrt{6})/2$
(B,2)	0	1	0	$(\sqrt{6}-1)/2$
(2,C)	1	0	0	0
(3,C)	0	0	1	0
(4,D)	0	1	0	$(\sqrt{6}-1)/2$
(1,2)	1	0	1	0
(2,3)	0	1	1	$(\sqrt{6}-1)/2$
(3,4)	0	1	0	$(\sqrt{6}-1)/2$
$(u(1),1^+)$	$(\tilde{v}_1^{C'})$ 1	$(\tilde{v}_1^{D'})$ 0	$(\tilde{v}_1^{C''})$ 1	$(\tilde{v}_1^{D''})$ 0
$(u(2),2^+)$	$(\tilde{v}_2^{C'})$ 0	$(\tilde{v}_2^{D'})$ 1	$(\tilde{v}_2^{C''})$ 0	$(\tilde{v}_2^{D''})$ $(\sqrt{6}-1)/2$
$(2^-,2^+)$	$(\bar{v}_2^{C'})$ 0	$(\bar{v}_2^{D'})$ 0	$(\bar{v}_2^{C''})$ 1	$(\bar{v}_2^{D''})$ 0
$(3^-,3^+)$	0	1	0	$(\sqrt{6}-1)/2$

Pour $\{v_a^{d'}\}$, on a $\quad \tilde{v}_1' = 1 \Rightarrow c_{u(1),1^+}(\underline{v}') = 4$

et $\quad \left.\begin{array}{l} \bar{v}_2' = 0 \\ \tilde{v}_2' = 1 \end{array}\right\} \Rightarrow \quad c_{u(2),2^+}(\underline{v}^*) = 4$

Pour $\{v_a^{d''}\}$, on a $\quad \tilde{v}_1'' = 1 \Rightarrow \quad c_{u(1),1^+}(\underline{v}'') = 4$

$$\left.\begin{array}{l} \bar{v}_2'' = 1 \\ \tilde{v}_2'' = (\sqrt{6}-1)/2 \end{array}\right\} \Rightarrow \quad c_{u(2),2^+}(\underline{v}'') = 6$$

On peut facilement calculer les longueurs des chemins les plus courts de A à C et de B à C pour chacune des affectations. On obtient :

$$\pi_A^C(\underline{v}') = \pi_A^C(\underline{v}'') = 14,$$

$$\pi_B^D(\underline{v}'') = 14,$$

et $$\pi_B^D(\underline{v}'') = 16.$$

On peut facilement vérifier que, dans les deux affectations, tous les usagers empruntent uniquement des arcs appartenant au sous-graphe des chemins les plus courts vers leur destination, et donc que ce sont deux affectations d'équilibre.

Pourtant, les usagers peuvent utiliser indifféremment le chemin {A-1-2-C} ou le chemin {A-1-2-3-C} pour se rendre de A à C, et ce dans n'importe quelle proportion. Qui plus est, à chaque répartition possible des usagers de la paire O-D A-C correspond une répartition différente des usagers de la paire B-D entre les chemins {B-D} et {B-2-3-4-D} et il existe des affectations pour lesquelles le coût de transport entre B et D est sensiblement différent (comme ici dans l'exemple pour $\{v_a^{d'}\}$ et $\{v_a^{d''}\}$). Peut-on alors vraiment parler d'équilibre? Comment déterminer la "bonne" répartition des usagers de la paire A-C? Le modèle ne fournit aucune réponse à ce genre de questions.

A la lumière de l'exemple, il devient clair que, s'il n'existe pas dans le modèle de mécanisme tendant à fixer les flots totaux d'usagers sur les sections de lignes, on risque fort d'aboutir à des situations

dans lesquelles on n'a pas d'équilibre véritable. C'est donc ici qu'interviennent les coûts variables sur les sections de lignes. L'utilisation de tels coûts va nous permettre d'éliminer une bonne partie des situations d'équilibres multiples; cependant, comme nous le verrons dans la discussion des propriétés du modèle au chapitre neuvième, cela n'est pas suffisant en général pour garantir l'existence d'un équilibre unique.

CHAPITRE CINQUIÈME

FONCTIONS D'ATTENTE ET QUASI-FILES D'ATTENTE

5.1 Introduction

Pour que le modèle présenté au chapitre quatrième soit complet, nous devons définir des fonctions nous donnant le temps d'attente des usagers aux arrêts. Pour obtenir ces fonctions, il existe plusieurs approches possibles : dérivation d'une forme fonctionnelle à partir de résultats empiriques ou de résultats de simulation, modèle théorique d'arrêt, spécification a priori d'une classe de fonctions ayant des propriétés désirables, etc. La méthodologie que nous avons retenue est la suivante : dans un premier temps, on considère que chaque arrêt constitue, de manière isolée, un système définissant un processus stochastique du type "file d'attente" (c'est ce que nous appelons une quasi-file d'attente — en anglais, Powell (1981) utilise le terme "scheduled departure queue") dont on peut déterminer de manière exacte les mesures de performance (longueur de la file d'attente, temps moyen d'attente) en régime stationnaire; par la suite, on peut utiliser des résultats exacts pour construire des formules d'approximation simples qui serviront à définir les fonctions d'attente du modèle de flot.

Notre méthodologie se reflète dans le plan de ce chapitre : après avoir précisé le concept de quasi-file d'attente et montré pourquoi un arrêt d'autobus peut être analysé comme tel, nous faisons d'abord un bref

survol des travaux antérieurs sur le sujet et nous présentons certains résultats exacts pour le cas le plus simple de quasi-files d'attente avec service en groupes (arrivées des "clients" — usagers — selon des processus de Poisson et intervalles de service — intervalles interautobus — qui suivent des lois exponentielles); nous tirons ensuite de ces résultats des bornes sur les temps d'attente et certaines formules d'approximation; par la suite, nous portons notre attention sur les quasi-files d'attente avec intervalles de service distribués selon des lois d'Erlang pour trouver une série de formules d'approximation pour les temps d'attente dans de tels systèmes; nous terminons en montrant comment la plus simple de ces approximations peut être exprimée par rapport aux variables du modèle du chapitre quatrième et en faisant une critique rapide des hypothèses sous-jacentes à notre méthodologie.

5.2 L'arrêt d'autobus en tant que quasi-file d'attente

Le phénomène qui se déroule à un arrêt particulier d'un système de transport en commun constitue de toute évidence un phénomène du type "file d'attente". En effet, les usagers se présentent à l'arrêt selon un certain processus et forment une file; quand un autobus passe à l'arrêt, un certain nombre d'usagers peuvent monter à bord dépendant du nombre de places libres; puis, d'autres usagers se présentent, un autre autobus arrive et ainsi de suite. Si l'on peut déterminer les distributions de probabilité régissant les arrivées des usagers, les intervalles entre les autobus et le nombre de places libres à bord de ceux-ci, il devient possible de modéliser le phénomène comme un processus stochastique.

A ce point-ci de notre analyse, il convient de souligner une particularité de ce système d'attente : dans un processus de file d'attente classique ("service queue" dans la terminologie de Powell, 1981), quand le serveur a fini de servir un groupe d'usagers et que la file d'attente est vide, il reste disponible jusqu'à l'arrivée du prochain usager qui sera servi dès qu'il se présentera (sans attendre, donc), alors qu'ici quand un autobus se présente à l'arrêt et que personne n'attend (ce qui est l'analogue de la situation précédente), il ne s'y arrête pas (et donc il n'est pas disponible), ce qui fait que, si un usager se présente, il devra attendre l'autobus suivant, exactement comme si un groupe d'usagers était monté à bord de l'autobus précédent. Ceci nous amène à conclure que le processus étudié appartient à la classe des quasi-files d'attente ("scheduled departure queue" dans la terminologie de Powell). Kadosch (1976) semble avoir été le premier à utiliser ce type de processus stochastiques pour le calcul de temps d'attente dans les transports en commun, alors que Gendreau (1977) a précisé le concept de quasi-file d'attente.

Si l'on suppose que les arrivées des passagers, les intervalles entre les autobus et le nombre de places libres à bord des autobus sont des variables aléatoires mutuellement indépendantes, le système d'attente constitué par un arrêt est une quasi-file d'attente avec service en groupes "ordinaire" par les correspondances suivantes :

- les arrivées des usagers à l'arrêt correspondent aux arrivées de "clients" dans la quasi-file;
- l'ensemble des autobus de la ligne desservant l'arrêt constituent le mécanisme de service;

- les intervalles entre les autobus correspondent aux intervalles de service;
- la taille des groupes pour le service est donnée par le nombre de places libres par autobus.

Dans la suite, nous utiliserons pour les différents types de quasi-files d'attente, la notation proposée par Gendreau, notation obtenue en ajoutant un préfixe "Q-" à la notation de Kendall (1953) utilisée pour les files d'attente classiques.

5.3 Survol des travaux antérieurs sur les quasi-files d'attente avec service en groupes

Bailey (1954) a été le premier à s'intéresser aux systèmes d'attente avec service en groupes. Le système qu'il a étudié correspond à une quasi-file d'attente $Q\text{-}M/G^{[K]}/1$ (arrivées selon un processus de Poisson, distribution générale par les intervalles de service, service en groupes de taille fixe). Par une technique de chaîne de Markov induite ("imbedded Markov chain"), Bailey a réussi à déterminer la transformée z de la longueur de la file d'attente aux instants de service; dans le cas où la distribution des intervalles de service est une Erlang (système $Q\text{-}M/E_k^{[K]}/1$) ou une loi déterministe (système $Q\text{-}M/D^{[K]}/1$), cette transformée peut être exprimée en fonction de certaines des racines d'une équation complexe qui varie selon les distributions impliquées.

En terminant, soulignons que, toujours dans les cas $Q\text{-}M/E_k^{[K]}/1$ et $Q\text{-}M/D^{[K]}/1$, Bailey a défini une expression donnant le temps moyen d'attente en régime stationnaire en fonction des racines de l'équation caractéristique associée au système.

Dans la foulée de Bailey, Downton (1955, 1956) a montré que la transformée de Laplace de la distribution du temps d'attente pouvait s'écrire simplement en fonction de certaines des racines de l'équation caractéristique associée au système pour les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$ et $Q\text{-}M/D^{[K]}/1$. Downton a aussi calculé le temps d'attente moyen pour certaines valeurs des paramètres de service; ces résultats nous seront particulièrement utiles dans la suite quand nous voudrons évaluer la "qualité" des formules d'approximation que nous proposons (section 5.7).

Dans leur article sur l'utilisation des distributions d'Erlang dans les systèmes d'attente, Kotiah, Thompson et Waugh (1969) ont traité rapidement la question des quasi-files d'attente avec service en groupes dans le cas où les distributions des intervalles de services sont des lois d'Erlang plus générales que celles considérées par Bailey et Downton Il est surprenant de voir que ces auteurs n'ont pas été frappés par le lien très étroit existant entre une quasi-file d'attente $Q\text{-}M/E_k^{[K]}/1$ et une file d'attente $E_K/E_k/1$ pour des paramètres identiques, alors que le plus gros de leur article porte justement sur les files $E_K/E_k/1$. C'est ce lien que nous exploitons dans la section 5.7 pour définir notre formule d'approximation de base pour les temps d'attente.

Petersen (1971) a repris les travaux de Bailey et Downton et fait l'extension de leurs résultats au cas du service en groupes de tailles aléatoires (quasi-files $Q\text{-}M/E_k^{[Y]}/1$ et $Q\text{-}M/D^{[Y]}/1$).

En 1976, Kadosch, qui semble avoir été totalement ignorant des travaux de Bailey, Downton et Petersen, a retrouvé des résultats semblables à ceux de ce dernier si la taille des groupes est donnée par une loi binomiale.

Gendreau (1977) a utilisé une approche directe (c'est-à-dire sans passer par les chaînes de Markov induites) pour calculer la distribution du nombre de personnes dans la file et la distribution du temps d'attente pour une quasi-file $Q\text{-}M/M^{[K]}/1$ en régime stationnaire. Les résultats obtenus sont semblables à ceux de Bailey pour le cas particulier des intervalles de service exponentiels.

Dans sa thèse de doctorat, Powell (1981) a étudié de manière très approfondie les quasi-files d'attente $Q\text{-}M^{[X]}/G^{[Y]}/1$ (c'est-à-dire avec des arrivées selon des processus de Poisson composés, des intervalles de service quelconques et des groupes de taille aléatoire) et certains systèmes qui leur sont apparentés (entre autres, systèmes avec annulation de certains services, systèmes avec priorités, systèmes avec intervalles de service corrélés). Il serait impossible de résumer en quelques lignes un travail aussi considérable; nous nous contentons donc ici d'encourager fortement le lecteur à consulter cette thèse. Une remarque importante : comme ses précédesseurs, Powell a obtenu des résultats qui s'expriment en fonction des racines des équations caractéristiques associées aux systèmes

qu'il étudie, ce qui a comme conséquence de les rendre extrêmement difficiles à utiliser dans le cadre d'un modèle global comme celui que nous avons proposé; c'est pour cette raison que nous allons devoir nous en remettre à des formules d'approximation dans la suite.

5.4 Résultats exacts pour les quasi-files d'attente Q-M/M$^{[K]}$/1

Dans cette section, nous allons faire un bref résumé des principaux résultats exacts relatifs aux quasi-files d'attente Q-M/M$^{[K]}$/1. Les principaux symboles que nous utiliserons ici, ainsi que dans les sections suivantes, sont :

- λ, taux d'arrivée des usagers;
- μ, fréquence de service (inverse de la moyenne des intervalles de service);
- K, taille (fixe) maximale des groupes pour le service;
- $\rho = \frac{\lambda}{K\mu}$, coefficient d'achalandage du système.

Propriété 5.1 (Bailey, 1954)

Une quasi-file d'attente Q-M/M$^{[K]}$/1 (ou Q-M/E$_k^{[K]}$/1) possède un état stationnaire si et seulement si son coefficient d'achalandage ρ est strictement inférieur à 1.

Propriété 5.2 (Bailey, 1954)

Dans une quasi-file d'attente Q-M/M$^{[K]}$/1 en régime stationnaire, les probabilités d'équilibre pour le nombre de personnes dans la file forment une série géométrique.

Propriété 5.3 (Gendreau, 1977), p.88)

Dans une quasi-file d'attente Q-M/$M^{[K]}$/1 en régime stationnaire, les probabilités d'équilibre pour le nombre de personnes dans la file, dénotées $\bar{p}_n$, n=0,1,..., sont données par :

$$\bar{p}_n = (1-r_0)\, r_0^n,\ n = 0,1,...$$

où r_0 est l'unique racine dans (0,1) de l'équation caractéristique

$$r^{K+1} - (\lambda+\mu)\, r + \lambda = 0 .$$

Propriété 5.4 (Bailey, 1954; Gendreau, 1977, p. 93)

Le temps d'attente moyen dans la file d'une quasi-file d'attente Q-M/$M^{[K]}$/1 en régime stationnaire, dénoté W_q, est donné par

$$W_q = \frac{1}{\lambda(z_K - 1)}$$

où z_K est l'unique racine à l'extérieur du disque-unité de l'équation caractéristique

$$z^K \{1 + \frac{\lambda}{\mu}(1-z)\} - 1 = 0$$

(formulation de Bailey).

W_q peut aussi s'exprimer sous la forme

$$W_q = \frac{r_0}{\lambda(1-r_0)} = \frac{1}{\mu(1-r_0^K)}$$

où r_0 est l'unique racine dans (0,1) de l'équation caractéristique

$$\mu r^{K+1} - (\lambda+\mu)r + \lambda = 0$$

(formulation de Gendreau).

Remarque : Il est facile de voir que, dans la proposition précédente, les deux formulations employées sont équivalentes; pour s'en assurer, il suffit de poser $z = 1/r$ dans la formulation de Bailey pour retrouver celle de Gendreau (on a alors $z_K = 1/r_0$).

Propriété 5.5 (Gendreau, 1977, p. 93)

Dans une quasi-file d'attente Q-M/$M^{[K]}$/1 en régime stationnaire, le temps d'attente dans la file suit une loi exponentielle.

5.5 Bornes et approximations pour le temps d'attente moyen dans les quasi-files d'attente Q-M/$M^{[K]}$/1 en régime stationnaire

Dans le cadre d'un modèle complexe, il est assez gênant que certaines fonctions ne soient pas exprimées sous une forme analytique; c'est donc dire que si nous voulons utiliser des quasi-files d'attente Q-M/$M^{[K]}$/1 pour représenter les arrêts d'autobus dans le modèle du chapitre quatrième, nous devrons trouver une manière d'exprimer le temps d'attente moyen dans la file sans avoir à résoudre d'équation caractéristique. Une solution possible consiste à borner W_q dans un premier temps, puis à l'approximer à partir des bornes obtenues dans un second; c'est ce que nous faisons dans cette section.

5.5.1 Bornes pour W_q dans les quasi-files Q-M/$M^{[K]}$/1

Nous avons vu précédemment que le temps d'attente moyen dans la

file d'une quasi-file d'attente Q-M/$M^{[K]}$/1 en régime stationnaire, W_q était donné par

$$W_q = \frac{r_0}{\lambda(1-r_0)} = \frac{1}{\mu(1-r_0^K)}$$

où r_0 est l'unique racine dans (0,1) de l'équation caractéristique

$$\mu\, r^{K+1} - (\lambda+\mu)r + \lambda = 0 \quad . \tag{1}$$

Pour borner W_q, on peut donc chercher des bornes sur r_0 (ou r_0^K), puis les substituer dans les expressions nous donnant W_q.

Remarquons tout d'abord que l'équation (1) peut être réécrite sous la forme

$$r^{K+1} - (K\rho+1)\, r + K\rho = 0 \tag{2}$$

pour une simple substitution $\frac{\lambda}{\mu} = K\rho$ et une division des deux membres de l'équation par μ. D'autre part, on peut noter que r=1 est une racine de l'équation. On obtient donc, après une mise en évidence du facteur (r-1),

$$(2) \iff (r-1)(r^K + r^{K-1} + \ldots + r^2 + r - K\rho) = 0$$

Puisque l'on cherche la racine de (2) appartenant à (0,1), on peut diviser par (r-1) et résoudre à la place

$$r^K + r^{K-1} + \ldots + r^2 + r - K\rho = 0 \tag{3}$$

$$\iff \sum_{i=1}^{K} r^i = K\rho \tag{4}$$

Notre problème peut donc se ramener à la recherche de bornes sur r_0 sachant que $r_0 \in (0,1)$ et que $\sum_{i=1}^{K} r_0^i = K\rho$. Pour ce faire, nous allons tout d'abord démontrer une série de lemmes qui nous seront fort utiles par la suite.

Lemme 5.1

Soit i,j et N des nombres entiers tels que $0 \leq i \leq j \leq \frac{N}{2}$, et soit $r \in (0,1)$, alors

$$(r^i + r^{N-i}) \geq (r^i + r^{N-j})$$

Démonstration

On a

$$(r^i + r^{N-i}) - (r^j + r^{N-j}) = (r^i - r^j) - (r^{N-j} - r^{N-i})$$

$$= r^i(1-r^{j-i}) - r^{N-j}(1 - r^{j-i})$$

$$(r^i + r^{N-i}) - (r^j + r^{N-j}) = r^i(1-r^{j-i})(1-r^{N-j-i})$$

Comme $i \leq j \leq \frac{N}{2}$, on a $N-j-i \geq 0$ et $j-i \geq 0$. Comme, d'autre part $r \in (0,1)$, on en conclut que

$$r^{N-j-1} \leq 1 \quad \text{et} \quad r^{j-i} \leq 1 ,$$

ce qui implique que

$$(r^i + r^{N-i}) - (r^j + r^{N-j}) \geq 0$$

$$\Leftrightarrow \qquad r^i + r^{N-i} \geq r^j + r^{N-j}$$

□

Lemme 5.2

Soit K un entier strictement positif et soit $r \in (0,1)$, alors

$$r + r^K \geq \frac{2}{K} \sum_{j=1}^{K} r^j$$

Démonstration

Pour K pair, on a

$$\sum_{j=1}^{K} r^j = \sum_{j=1}^{K/2} (r^j + r^{K+1-j}) \leq \sum_{j=1}^{K/2} (r^1 + r^K),$$

en appliquant le lemme 5.1 avec N=K+1 et i=1, pour j=1,2,...,K/2.

On en déduit que

$$\frac{K}{2} (r^1 + r^K) \geq \sum_{j=1}^{K} r^j$$

$$\Leftrightarrow \quad r + r^K \geq \frac{2}{K} \sum_{j=1}^{K} r^j$$

Pour K impair, on a

$$\sum_{j=1}^{K} r_j = \sum_{j=1}^{(K-1)/2} (r^j + r^{K+1-j}) + r^{(\frac{K+1}{2})} \leq \sum_{j=1}^{(K-1)/2} (r^1 + r^K) + r^{(\frac{K+1}{2})}$$

en appliquant le lemme 5.1 avec N=K+1 et i=1, pour j=1,2,...,(K-1)/2.

D'autre part, toujours par le lemme 5.1,

$$2r^{(\frac{K+1}{2})} \leq (r^1 + r^K) \Rightarrow r^{(\frac{K+1}{2})} \leq \frac{r^1 + r^K}{2} \quad .$$

De tout ceci, on déduit que

$$\frac{(K-1)}{2}(r^1+r^K)+(\frac{r^1+r^K}{2}) \geq \sum_{j=1}^{K} r^j$$

$$\Leftrightarrow \qquad r+r^K \geq \frac{2}{K}\sum_{j=1}^{K} r^j \quad .$$

La propriété est donc démontrée quelque que soit la parité de K.

□

Lemme 5.3

Soit K un entier strictement positif et soit $r\in(0,1)$, alors

$$r^{K+1} \leq (\frac{1}{K}\sum_{i=1}^{K} r_i)^2 \quad .$$

Démonstration

Pour K pair, on a

$$\sum_{i=1}^{K} r^i = \sum_{i=1}^{K/2} (r^i + r^{K+1-i}) \geq \sum_{i=1}^{K/2} (r^{K/2} + r^{K/2+1}),$$

en appliquant le lemme 5.1 avec N=K+1 et j=K/2, pour i=1,2,...,K/2.

On en déduit que

$$\frac{K}{2}(r^{K/2}+r^{K/2+1}) \leq \sum_{i=1}^{K} r^i$$

$$\Leftrightarrow \qquad r^{K/2}+r^{K/2+1} \leq \frac{2}{K}\sum_{i=1}^{K} r^i \quad .$$

Or, pour $a,b,c > 0$,

$$b+c = 2a \Rightarrow bc \leq a^2 ,$$

et, a fortiori,

$$b+c \leq 2a \Rightarrow bc \leq a^2 .$$

On peut donc conclure que

$$r^{K/2}+r^{K/2+1} \leq \left(\frac{\sum_{i=1}^{K} r^i}{K} \right)^2$$

$$\Leftrightarrow \quad r^{K+1} \leq \left(\frac{1}{K} \sum_{i=1}^{K} r^i \right)^2 .$$

Pour <u>K impair</u>, on a

$$\sum_{i=1}^{K} r^i = \sum_{i=1}^{(K-1)/2} (r^i+r^{K+1-i})+r^{(\frac{K+1}{2})} \geq \sum_{i=1}^{(K-1)/2} (r^{(\frac{K+1}{2})} + r^{(\frac{K+1}{2})})+r^{(\frac{K+1}{2})}$$

en appliquant le lemme 5.1 avec $N=K+1$ et $j=(K+1)/2$, pour $i=1,2,\ldots,(K-1)/2$.

On en déduit que

$$\frac{(K-1)}{2} (r^{(\frac{K+1}{2})} + r^{(\frac{K+1}{2})}) + r^{(\frac{K+1}{2})} \leq \sum_{i=1}^{K} r^i$$

$$\Leftrightarrow \quad K\, r^{(\frac{K+1}{2})} \leq \sum_{i=1}^{K} r^i$$

$$\Leftrightarrow \quad r^{(\frac{K+1}{2})} \leq \frac{1}{K} \sum_{i=1}^{K} r^i$$

$$\Rightarrow \quad r^{K+1} \leq (\frac{1}{K} \sum_{i=1}^{K} r^i)^2$$

La propriété est donc démontrée quelle que soit la parité de K.

□

Nous sommes maintenant en mesure de déterminer des bornes inférieure et supérieure sur la valeur de r_0.

Théorème 5.1

Soit K un entier strictement positif et soit $\rho\in(0,1)$, alors la racine unique r_0 de l'équation

$$r^{K+1} - (K\rho+1)\, r + K\rho = 0 \qquad (2)$$

dans l'intervalle (0,1) satisfait les inégalités

$$\frac{K\rho}{K\rho + \frac{(1-\rho)}{K\rho+1}\,[2K\rho + (1-\rho)]} \leq r_0 \leq \frac{K\rho+\rho^2}{K\rho+1}$$

Démonstration

Démontrons tout d'abord que $r_0 \leq \frac{K\rho+\rho^2}{K\rho+1}$.

On a vu auparavant que si r_0 était la racine dans (0,1) de l'équation (2), c'était aussi la racine dans (0,1) de l'équation (4)

$$\sum_{i=1}^{K} r^i = K\rho .$$

On a donc

$$\sum_{i=1}^{K} r_0^i = K\rho .$$

Par le lemme 5.3, on peut alors dire que

$$r_0^{K+1} \leq (\frac{1}{K} \sum_{i=1}^{K} r_0^i)^2 = (\frac{1}{K}(K\rho))^2 = \rho^2 \quad . \tag{5}$$

D'autre part, r_0 étant solution de (2), on a

$$r_0^{K+1} - (K\rho+1)r_0 + K\rho = 0$$

$$\Leftrightarrow \qquad r_0 = \frac{K\rho+r_0^{K+1}}{K\rho+1} \tag{6}$$

En combinant (5) et (6), on obtient

$$r_0 \leq \frac{K\rho+\rho^2}{K\rho+1} \quad ,$$

ce qui est bien le résultat désiré.

Pour ce qui est de la borne inférieure, on a que

$$\sum_{i=1}^{K} r_0^i = K\rho$$

$$\Rightarrow \qquad r_0 + r_0^K \geq 2\rho \quad \text{par le lemme 5.2.}$$

De ceci, on tire

$$(7) \qquad r_0^K \geq 2\rho - \frac{K\rho+\rho^2}{K\rho+1} \quad , \quad \text{car} \quad r_0 \leq \frac{K\rho+\rho^2}{K\rho+1} \, .$$

D'autre part,

$$r_0^{K+1} - (K\rho+1)\, r_0 + K\rho = 0$$

$$\Leftrightarrow \qquad r_0(K\rho+1 - r_0^K) = K\rho$$

$$\Leftrightarrow \qquad r_0 = \frac{K\rho}{K\rho+1-r_0^K} \tag{8}$$

En combinant (7) et (8), on obtient

$$r_0 \geq \frac{K\rho}{K\rho+1 - [2\rho - \frac{K\rho+\rho^2}{K\rho+1}]}$$

$$\Leftrightarrow \quad r_0 \geq \frac{K\rho}{K\rho + [\frac{K\rho+1-2K\rho^2-2\rho+K\rho+\rho^2}{K\rho+1}]}$$

$$\Leftrightarrow \quad r_0 \geq \frac{K\rho}{K\rho + \frac{(1-\rho)}{K\rho+1}[2K\rho+(1-\rho)]}$$

ce qui est bien le résultat que l'on cherchait.

□

Corollaire 5.1

Soit K un entier strictement positif et soit $\rho\in(0,1)$, alors la racine unique de l'équation

$$r^{K+1} - (K\rho+1)\, r + K\rho = 0$$

dans l'intervalle (0,1) satisfait les inégalités

$$r_0 > \frac{K\rho}{K\rho+2(1-\rho)}$$

et $$(1-r_0^K) \leq \frac{(1-\rho)}{K\rho+1}[2K\rho+(1-\rho)] \ .$$

Démonstration

Par le théorème 5.1, on a

$$r_0 \geq \frac{K\rho}{K\rho + \frac{(1-\rho)}{K\rho+1}[2K\rho+(1-\rho)]} > \frac{K\rho}{K\rho + \frac{(1-\rho)}{K\rho+1}[2K\rho+2]} = \frac{K\rho}{K\rho+2(1-\rho)} \ .$$

D'autre part, puisque

$$r_0^K \geq 2\rho - \frac{K\rho+\rho^2}{K\rho+1} \quad ,$$

on déduit facilement des calculs de la seconde partie de la démonstration du théorème 5.1 que

$$1 - r_0^K \leq \frac{(1-\rho)}{K\rho+1} \; [2K\rho+(1-\rho)] \, .$$

□

Connaissant des bornes sur r_0, nous sommes maintenant en mesure de borner W_q.

Théorème 5.2

Le temps d'attente moyen dans la file d'une quasi-file d'attente Q-M/M$^{[K]}$/1 en régime stationnaire, W_q, satisfait les inégalités suivantes :

$$\frac{1}{\mu(1-\rho)} \cdot \frac{K\rho+1}{2K\rho+(1-\rho)} \leq W_q \leq \frac{1}{\mu(1-\rho)} \cdot \frac{K+\rho}{K(1+\rho)} \quad .$$

Démonstration

On sait que W_q est donné par les relations

$$W_q = \frac{r_0}{(1-r_0)} = \frac{1}{(1-r_0^K)}$$

où r_0 est l'unique racine dans (0,1) de l'équation

$$r^{K+1} - (K\rho+1)r + K\rho = 0.$$

On sait aussi par le corollaire 5.1 que

$$(1-r_0^K) \leq \frac{(1-\rho)}{K\rho+1}\,[\,2K\rho+(1-\rho)\,]\ .$$

De ceci, on conclut que

$$W_q = \frac{1}{\mu(1-r_0^K)} \geq \frac{1}{\mu} \cdot \frac{K\rho+1}{(1-\rho)[\,2K\rho+(1-\rho)]} \qquad (9)$$

D'autre part, on a, par le théorème 5.1, que

$$r_0 \leq \frac{K\rho+\rho^2}{K\rho+1}\ .$$

Ceci implique que

$$W_q = \frac{r_0}{\lambda(1-r_0)} \leq \frac{\frac{K\rho+\rho^2}{K\rho+1}}{\lambda(1-\frac{K\rho+\rho^2}{K\rho+1})}$$

$$\Rightarrow \quad W_q \leq \frac{K\rho+\rho^2}{\lambda(1-\rho^2)} = \frac{\rho}{\lambda(1-\rho)} \cdot \frac{K+\rho}{1+\rho}$$

Après la substitution $\frac{\rho}{\lambda} = \frac{1}{K\mu}$ on obtient

$$W_q \leq \frac{1}{K\mu(1-\rho)} \cdot \frac{K+\rho}{K(1+\rho)} \qquad (10)$$

Réécrivant (9) et (10), on obtient finalement

$$\frac{1}{\mu(1-\rho)} \cdot \frac{K\rho+1}{2K\rho+(1-\rho)} \leq W_q \leq \frac{1}{\mu(1-\rho)} \cdot \frac{K+\rho}{K(1+\rho)}$$

□

On peut remarquer que les deux bornes sur W_q sont exprimées de manière à mettre en évidence un facteur $\frac{1}{\mu(1-\rho)}$. Cette valeur correspond au temps d'attente moyen dans la file d'une quasi-file Q-M/M/1 avec fréquence de service μ et coefficient d'achalandage ρ. D'une certaine façon, $\frac{1}{\mu(1-\rho)}$ constitue un estimé "naïf" de W_q et les bornes obtenues permettent d'estimer l'écart existant entre cet estimé et la véritable valeur de W_q.

Pour la suite, posons

$$\beta^-(K) = \frac{K\rho+1}{2K\rho+(1-\rho)} \quad \text{et} \quad \beta^+(K) = \frac{K+\rho}{K(1+\rho)} \quad .$$

Le résultat du théorème 5.2 peut alors être réécrit sous la forme

$$\frac{1}{\mu(1-\rho)}\,\beta^-(K) \le W_q \le \frac{1}{\mu(1-\rho)}\,\beta^+(K) \quad .$$

On peut facilement vérifier les propriété suivantes :

- $\beta^-(1) = \beta^+(1) = 1$
- $\frac{1}{2} < \beta^-(K) < \beta^+(K) < 1 \,, \qquad \forall\, K>1, \qquad \forall\, \rho\in(0,1)$
- $\lim_{K\to\infty} \beta^-(K) = \frac{1}{2}$
- $\lim_{K\to\infty} \beta^+(K) = \frac{1}{1+\rho}$
- $\lim_{K\to\infty} \frac{\beta^-(K)}{\beta^+(K)} = \frac{1+\rho}{2}$
- $\frac{\beta^-(K)}{\beta^+(K)} > \frac{1+\rho}{2} \,, \qquad \forall\, K\ge 1 \,, \qquad \forall\, \rho\in(0\ 1)$

Ces quelques propriété nous permettent de voir que nos bornes sont

exactes pour K = 1 et qu'elles s'écartent progressivement l'une de l'autre quand K augmente. On note aussi que les bornes sont beaucoup plus serrées, en termes relatifs, quand le coefficient d'achalandage est proche de 1. Par exemple, si on a $\rho = 0{,}9$, le rapport $(1+\rho)/2$ est égal à 0,95, ce qui fait qu'en choisissant n'importe quelle valeur entre $\beta^-(K)$ et $\beta^+(K)$ pour estimer W_q, on est assuré que l'erreur relative sera au plus de l'ordre de 5%.

A titre indicatif, nous avons tabulé les valeurs des bornes inférieure et supérieure sur W_q pour une quasi-file d'attente $Q\text{-}M/M^{[K]}/1$ avec $\lambda = 8$ et $\rho = 0{,}8$ quand K prend les valeurs 1, 10, 50 et 100.

K	$1/\mu$	Borne inférieure	Borne supérieure
1	0,1	0,50	0,50
10	1,0	2,78	3,00
50	5,0	12,78	14,11
100	10,0	25,28	28,00

Tableau 5.1 : Bornes sur W_q pour une quasi file d'attente $Q\text{-}M/M^{[K]}/1$ avec $\lambda = 8$ et $\rho = 0{,}8$

5.5.2 Formules d'approximation pour W_q dans les quasi-files d'attentes $Q\text{-}M/M^{[K]}/1$

A partir des bornes que nous venons d'établir sur W_q, il est relativement facile de suggérer des formules d'approximation pour W_q. Comme $\frac{1}{\mu(1-\rho)} \cdot \beta^-(K) \leq W_q \leq \frac{1}{\mu(1-\rho)} \beta^+(K)$, nous allons étudier les formules de la forme $W_q \simeq \frac{1}{\mu(1-\rho)} \beta(K)$ où $\beta(K)$ reste à déterminer.

La première définition possible pour $\beta(K)$ est la moyenne arithmétique de $\beta^-(K)$ et $\beta^+(K)$, c'est-à-dire $\bar{\beta}(K) = \frac{\beta^-(K) + \beta^+(K)}{2}$. Malheureusement, on a

$$\bar{\beta}(K) = \frac{1}{2}\left[\frac{K\rho+1}{2K\rho+(1-\rho)} + \frac{K+\rho}{K(1+\rho)}\right] \quad ,$$

ce qui n'est pas l'expression la plus intéressante qui puisse être.

Une autre solution possible consiste à additionner terme à terme les numérateurs et les dénominateurs de $\beta^-(K)$ et $\beta^+(K)$. Après quelques tâtonnements, on obtient une formule assez satisfaisante si l'on fait précéder cette opération par une multiplication du numérateur et du dénominateur de $\beta^-(K)$ par $(1+\rho)$ et par une multiplication de ceux de $\beta^+(K)$ par 2ρ. On a alors

$$\beta^-(K) = \frac{(1+\rho)(K\rho+1)}{(1+\rho)[\,2K\rho+(1-\rho)]} \leq \frac{(1+\rho)(K\rho+1)+2\rho(K+\rho)}{(1+\rho)[\,2K\rho+(1-\rho)]+2\rho[\,K(1+\rho)]} \leq \frac{2\rho(K+\rho)}{2\rho[\,K(1+\rho)]} = \beta^+(K)$$

Si l'on pose

$$\tilde{\beta}(K) = \frac{(1+\rho)(K\rho+1)+2\rho(K+\rho)}{(1+\rho)[\,2K\rho+(1-\rho)]+2\rho[\,K(1+\rho)]} \quad ,$$

on obtient après avoir regroupé les différents termes

$$\tilde{\beta}(K) = \frac{K\rho(3+\rho)+(1+\rho+2\rho^2)}{(1+\rho)[\,4K\rho+(1-\rho)]} \quad .$$

On peut remarquer que

$$\tilde{\beta}(1) = \beta^-(1) = \beta^+(1) = 1$$

et que $\lim_{K\to\infty} \tilde{\beta}(K) = \lim_{K\to\infty} \left(\frac{\beta^-(K)+\beta^+(K)}{2}\right) = \frac{3+\rho}{4(1+\rho)}$.

A la lumière de ces deux résultats, il semble donc que $W_q \simeq \frac{1}{\mu(1-\rho)} \tilde{\beta}(K)$ constitue une formule d'approximation relativement satisfaisante. Il faut néanmoins se rendre compte qu'il n'est guère plus facile de travailler avec $\tilde{\beta}(K)$ qu'avec $\bar{\beta}(K)$. Nous avons essayé de trouver des combinaisons de $\beta^-(K)$ et $\beta^+(K)$ qui puissent s'exprimer sous une forme plus simple que $\bar{\beta}(K)$ et $\tilde{\beta}(K)$, tout en satisfaisant les propriétés $\beta(1) = 1$ et $\lim_{K\to\infty} \beta(K) = \frac{3+\rho}{4(1+\rho)}$, mais nos efforts sont demeurés vains.

Le tableau 5.2 qui suit donne les valeurs de $\beta^-(K)$, $\beta^+(K)$, $\bar{\beta}(K)$ et $\tilde{\beta}(K)$ pour quelques valeurs possibles de ρ et de K.

ρ	K	$\beta^-(K)$	$\bar{\beta}(K)$	$\tilde{\beta}(K)$	$\beta^+(K)$
	2	0,765	0,771	0,771	0,778
	5	0,610	0,627	0,627	0,644
0,8	10	0,556	0,578	0,578	0,600
	50	0,511	0,437	0,538	0,564
	100	0,506	0,533	0,533	0,560
	2	0,800	0,816	0,815	0,833
	5	0,636	0,684	0,683	0,733
0,5	10	0,571	0,635	0,634	0,700
	50	0,515	0,594	0,594	0,673
	100	0,507	0,588	0,589	0,670

Tableau 5.2 : Valeurs de $\beta^-(K)$, $\beta^+(K)$, $\bar{\beta}(K)$ et $\tilde{\beta}(K)$ pour ρ=0,8 et 0,5 et K=2,5,10,50,100

5.6 Extension des résultats de Pollaczek-Khinchine aux quasi-files d'attente Q-M/G/1

5.6.1 Introduction

Dans la section précédente, nous avons défini une formule d'approximation pour le temps d'attente dans les quasi-files d'attente Q-M/$M^{[K]}$/1 dans laquelle on peut isoler un facteur correspondant au temps d'attente moyen dans une quasi-file Q-M/M/1 ayant même fréquence de service et même coefficient d'achalandage. Dans la même veine, on pourrait essayer d'établir un lien entre les temps d'attente dans les quasi-files Q-M/$G^{[K]}$/1 et les temps d'attente dans les quasi-files Q-M/G/1. Il faudrait alors dans un premier temps pouvoir calculer le temps d'attente moyen dans une quasi-file Q-M/G/1. Bien que nous ayons choisi, comme nous le verrons dans la section 5.7, d'utiliser une approche différente de celle que nous venons de suggérer pour la détermination des temps d'attente des quasi-files Q-M/$E_k^{[K]}$/1, il nous a paru intéressant d'inclure tout de même ici les résultats relatifs aux temps d'attente dans les quasi-files Q-M/G/1.

5.6.2 Rappel des résultats de Pollaczek-Khinchine pour les files d'attente classiques M/G/1

Les équations de Pollaczek-Khinchine nous permettent, avec l'aide des formules de Little, de déterminer toutes les mesures de performance habituelles pour les files d'attente M/G/1 : temps moyen passé dans le système, temps moyen d'attente dans la file, nombre moyen de personnes dans le système et longueur moyenne de la file. Pour notre part, nous allons nous restreindre à l'étude du temps d'attente moyen dans la file.

Le résultat de Pollaczek-Khinchine pour le temps d'attente (W_q) est généralement exprimé sous la forme suivante :

$$W_q = \frac{\rho^2+\lambda^2\sigma_s^2}{2\lambda(1-\rho)}$$

où λ est le taux d'arrivée des clients,

$1/\mu$ est la longueur moyenne d'un intervalle de service,

σ_s^2 est la variance de la longueur des intervalles de service

et $\rho(\equiv\frac{\lambda}{\mu})$ est le coefficient d'achalandage du système.

Le résultat peut être réécrit sous une forme quelque peu différente en effectuant une suite de manipulations algébriques simples :

$$W_q = \frac{\rho^2+\lambda^2\sigma_s^2}{2\lambda(1-\rho)}$$

$$= \frac{(\lambda/\mu)^2+\lambda^2\sigma_s^2}{2\lambda(1-\rho)}$$

$$= \frac{\lambda}{\mu}\,\frac{(1/\mu)^2+\sigma_s^2}{2(1/\mu)(1-\rho)}$$

$$W_q = \frac{\rho}{1-\rho}\cdot\frac{(1/\mu)^2+\sigma_s^2}{2(1/\mu)} \qquad (11)$$

Appelons S, la variable aléatoire "longueur d'un intervalle de service". On a alors

$$E[S] \equiv 1/\mu$$

et $$\text{Var } S \equiv \sigma_s^2 .$$

L'équation (11) devient

$$W_q = \frac{\rho}{1-\rho} \cdot \frac{E[S]^2 + \text{Var } S}{2E[S]}$$

$$\text{ou} \quad W_q = \frac{\rho}{1-\rho} \frac{E[S^2]}{2E[S]}$$

par l'identité $E[S^2] = E[S]^2 + \text{Var } S$.

Les termes de la forme $\frac{E[S^2]}{2E[S]}$ sont bien connus en théorie du renouvellement; en fait, $\frac{E[S^2]}{2E[S]}$ est la vie résiduelle moyenne pour un objet dont la vie utile est distribuée comme S. C'est donc dire que si le serveur est occupé quand un client se présente, il se passera en moyenne $\frac{E[S^2]}{2E[S]}$ unités de temps avant la fin du présent intervalle de service.

D'autre part, dans une file M/G/1 en régime stationnaire, la probabilité que le serveur soit occupé à un instant quelconque est égale au coefficient d'achalandage. C'est donc dire que $\rho \frac{E[S^2]}{2E[S]}$ correspond à l'espérance de l'intervalle séparant l'arrivée d'un nouveau client du moment auquel le serveur pourra débuter un nouvel intervalle de service (si le serveur est libre — ce qui se passe avec probabilité $(1-\rho)$ — la longueur de cet intervalle est nulle).

Si l'on pose $W_0 = \rho \frac{E[S^2]}{2E[S]}$, on aboutit à la relation

$$W_q = \frac{W_0}{1-\rho} \tag{12}$$

Cette forme nous permet d'interpréter le temps moyen d'attente comme le produit d'un temps d'attente initial (W_0) et d'un facteur de congestion du système ($\frac{1}{1-\rho}$).

5.6.3 Extension aux quasi-files d'attente Q-M/G/1

Dans cette sous-section, nous allons montrer que le résultat de Pollaczek-Khinchine s'applique aux quasi-files d'attente Q-M/G/1 à condition que l'on l'exprime sous la forme (12) (et nous sous la forme (11)).

Pour ce faire, considérons la chaîne de Markov induite quand on observe la quasi-file d'attente à la fin des intervalles de service (ou au début de ces intervalles puisque, dans une quasi-file d'attente, ces instants se confondent). Comme variable d'état, nous choisirons le nombre de personnes restant dans la file aux instants auxquels nous l'observons. Comme Bailey (1954) le fait remarquer, la matrice de transition que l'on obtient pour cette chaîne de Markov est identique à celle que l'on a pour la chaîne de Markov induite pour une file d'attente M/G/1 (à condition, bien sûr, que l'on ait les mêmes paramètres et les mêmes distributions de service pour les deux systèmes). On peut par conséquent utiliser directement les résultats découlant de l'analyse de la chaîne de Markov induite pour les files d'attente M/G/1, entre autres (Kleinrock, 1975, p. 187),

$$\overline{N} = \rho + \frac{\lambda^2 E[S^2]}{2(1-\rho)}$$

où $\overline{N}$ est le nombre moyen de personnes demeurant dans le système (dans la file à la fin d'un intervalle de service.

Il est maintenant extrêmement important de remarquer que, dans une quasi-file d'attente, les intervalles de service ne correspondent pas tous à des intervalles dans lesquels on sert effectivement un client. En fait, si l'on observe le système pendant une très longue période T, on aura en

moyenne λT arrivées de clients et μT intervalles de service en tout; c'est donc dire qu'une proportion ρ ($=\frac{\lambda T}{\mu T}$) des intervalles de service correspondent à des services réels aux clients et qu'une proportion $(1-\rho)$ de ces intervalles sont des intervalles de service "fictif". On peut donc écrire

$$\overline{N} = E[\text{\# clients dans le système à la fin d'un intervalle}]$$

$$\Leftrightarrow \quad \overline{N} = E[\text{\# clients dans le système à la fin d'un intervalle} \mid \text{service réel}]\ Pr\{\text{service réel}\} + E[\text{\# clients dans le système à la fin d'un intervalle} \mid \text{service fictif}]\ Pr\{\text{service fictif}\}$$

Si on appelle $\overline{N}_r$ et $\overline{N}_f$ les deux espérances conditionnelles, on obtient

$$\overline{N} = \overline{N}_r\ Pr\{\text{service réel}\} + \overline{N}_f\ Pr\{\text{service fictif}\}$$

$$\Leftrightarrow \quad \overline{N} = \overline{N}_r\ \rho + \overline{N}_f\ (1-\rho)$$

On peut trouver $\overline{N}_f$ facilement. Le système étant vide au début d'un intervalle de service fictif, les seuls clients qui s'y trouveront à la fin de cet intervalle, seront ceux arrivés pendant celui-ci. Or, le nombre moyen de clients se présentant durant un intervalle de service est tout simplement égal à λ (taux d'arrivée) multiplié par $1/\mu$ (longueur moyenne de l'intervalle). C'est donc dire que $\overline{N}_f = \lambda\ (1/\mu) = \rho$. Et alors

$$\overline{N} = \overline{N}_r\ \rho + \rho(1-\rho)$$

$$\Leftrightarrow \quad \rho + \frac{\lambda^2\ E[S^2]}{2(1-\rho)} = \overline{N}_r\ \rho + \rho(1-\rho)$$

$$\Leftrightarrow \quad \overline{N}_r = \rho + \frac{\lambda^2\ E[S^2]}{2\rho(1-\rho)}$$

Remarquons ici que $\overline{N}_r$ représente le nombre moyen de personnes qu'un client laisse derrière lui dans le système; mais, comme dans une quasi-file d'attente les changements d'états sont d'amplitude 1, on peut aussi affirmer que $\overline{N}_r$ correspond au nombre moyen d'usagers qu'un client trouve dans le système à son arrivée (Kleinrock, 1975, p. 176). D'autre part, les arrivées suivant un processus de Poisson, on sait que le nombre moyen d'usagers qu'un client trouve à son arrivée dans le système est égal en nombre moyen d'usagers dans le système, L. On a donc

$$L = \rho + \frac{\lambda^2 E[S^2]}{2\rho(1-\rho)} \quad .$$

Dans un système à serveur unique, le nombre moyen dans la file d'attente est $L_q = L-\rho$, ce qui fait que

$$L_q = \frac{\lambda^2 E[S^2]}{2\rho(1-\rho)} \quad .$$

Pour trouver W_q, il suffit maintenant d'appliquer la formule de Little $W_q = L_q/\lambda$. On obtient

$$W_q = \frac{\lambda E[S^2]}{2\rho(1-\rho)}$$

$$= \frac{\lambda E[S^2]}{2(\frac{\lambda}{\mu})(1-\rho)}$$

$$W_q = \frac{E[S^2]}{2 E[S](1-\rho)} \quad .$$

Dans le cas d'une quasi-file d'attente, on sait qu'en arrivant, un client ne trouvera jamais le serveur libre et donc $\frac{E[S^2]}{2 E[S]}$ correspond à l'expérance de la longueur de l'intervalle séparant l'arrivée du client

de la fin de l'intervalle de service en cours, c'est-à-dire à W_0. On peut donc conclure que

$$W_q = \frac{W_0}{1-\rho} \quad ,$$

ce qui constitue l'extention du résultat de Pollaczek-Khinchine aux quasi-files d'attente Q-M/G/1.

5.6.4 Interprétation intuitive

Par comparaison des cycles d'activité d'une file d'attente M/G/1 et d'une quasi-file Q-M/G/1 ayant des processus d'arrivée identiques et des intervalles de service réels identiques, il est possible d'en arriver à une interprétation "intuitive" intéressante du résultat précédent.

Quand un client arrive dans une file d'attente classique vide, le serveur s'en occupe immédiatement, alors que dans une quasi-file d'attente "semblable", le client doit attendre la fin d'un intervalle de service "fictif" avant d'être servi, ce qui implique un "retard" de longueur moyenne $\frac{E[S^2]}{2E[S]}$. Le serveur ayant pris ce "retard", tous les clients arrivant dans le même cycle d'activité seront retardés d'autant. Bien que les cycles d'activité des deux systèmes ne correspondent pas nécessairement exactement, globalement on obtient exactement l'effet suggéré. En effet,

$$E[\text{temps d'attente dans la quasi-file}] = \frac{E[S^2]}{2E[S](1-\rho)}$$

$$= \frac{\rho E[S^2] \quad (1-\rho)E[S^2]}{2E[S](1-\rho)}$$

$$= \frac{\rho E[S^2]}{(1-\rho)2E[S]} + \frac{E[S^2]}{2E[S]}$$

$$E[\text{temps d'attente dans la quasi-file}] = E[\text{temps d'attente dans la file}] + E[\text{retard}]$$

5.6.5 Lien avec les travaux de Finch

Finch (1959) a étudié une classe de systèmes d'attente dans lesquels un client qui arrive dans un système vide n'est pas servi immédiatement. Il a montré que, sous certaines conditions, on pouvait généraliser la formule de Pollaczek-Khinchine donnant la fonction caractéristique de la distribution du temps d'attente en régime stationnaire. Les résultats de cette section constituent donc en quelque sorte un cas particulier des résultats de Finch. Soulignons cependant que l'approche que nous avons employée est complètement différente de celle de Finch.

5.7 Temps d'attente moyen dans les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$

5.7.1 Introduction

Dans la sous-section 3.8.3, nous avions évoqué le problème du choix d'une famille de distributions adéquates pour les intervalles inter-autobus et suggéré d'utiliser les distributions d'Erlang. Au risque de nous répéter, nous croyons qu'il est essentiel de souligner que la distribution exponentielle et la distribution déterministe (constante) constituent des cas trop extrêmes (coefficient de variation égal à 1 dans le premier cas et à 0 dans le second) pour que l'on puisse s'en contenter dans le cadre d'un modèle d'affectation des passagers. Il faut donc trouver mieux. La solution la plus simple, à notre sens, consiste à utiliser l'Erlang et ce pour les deux raisons suivantes :

a) par un choix adéquat du paramètre de forme de l'Erlang, on peut ajuster la distribution théorique du modèle à une distribution empirique quelconque sur la base des deux premiers moments (à la condition que le coefficient de variation ne dépasse pas 1, ce qui devrait être vérifié en pratique);

b) nous disposons des résultats de tous les travaux antérieurs sur les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$.

Comme dans la section 5.5, notre but ici est l'obtention d'une formule d'approximation pour le temps d'attente moyen quand le système que nous étudions est en régime stationnaire. Pour les quasi-files $Q\text{-}M/M^{[K]}/1$, nous avons défini une formule d'approximation à partir de bornes sur l'unique racine d'intérêt de l'équation caractéristique associée au système; cette approche ne pourra pas être utilisée pour les quasi-files $Q\text{-}M/E_k^{[K]}/1$, car le temps d'attente moyen s'exprime en fonction d'un sous-ensemble de k racines de l'équation caractéristique. Pour contourner cette difficulté, nous allons tout d'abord montrer qu'il existe une relation directe entre les temps d'attente dans les quasi-files $Q\text{-}M/E_k^{[K]}/1$ et les temps d'attente dans les files d'attente classiques $E_K/E_k/1$; par la suite, nous allons construire une série d'approximations pour le temps d'attente dans les files $E_K/E_k/1$, série à laquelle correspondra une série "semblable" d'approximations pour le temps d'attente dans les quasi-files $Q\text{-}M/E_k^{[K]}/1$.

5.7.2 Rappel des résultats de Bailey et Downton

Dans cette sous-section, nous nous bornons à énoncer (sans les démontrer) deux théorèmes qui découlent directement des travaux de Bailey et Downton.

Afin d'alléger quelque peu la notation dans la suite de cette section, nous supposerons toujours implicitement ou explicitement que les unités de temps des systèmes étudiés ont été normalisées pour rendre

le taux moyen d'arrivée des clients égal à 1 (c'est-à-dire $\lambda = 1$).

Théorème 5.3 (Bailey, 1954)

Soit ρ (<1) le coefficient d'achalandage d'une quasi-file d'attente $Q\text{-}M/E_k^{[K]}/1$ en régime stationnaire, alors le temps moyen d'attente dans la file $W_q^{(K)}$, exprimé en unités de temps normalisées ($\lambda = 1$), est donnée par :

$$W_q^{(K)} = \sum_{j=K}^{K+k-1} (z_j - 1)^{-1} - K\rho \left(\frac{k-1}{2k}\right) \qquad (13)$$

où les z_j ($j = K, \ldots, K+k-1$) sont les k racines à l'extérieur du disque-unité de l'équation caractéristique

$$z^K [\, 1 + K\rho(1-z)/k \,]^k - 1 = 0 \qquad (14)$$

Quand on laisse tendre le paramètre de forme k de la famille des distributions d'Erlang vers l'infini, on obtient une loi déterministe. Bailey avait aussi considéré ce type de distribution dans son article, mais, pour des raisons de clarté, nous avons préféré ici reprendre directement les formules données par Downton.

Théorème 5.4 (Downton, 1955)

Soit ρ (<1) le coefficient d'achalandage d'une quasi-file d'attente $Q\text{-}M/D^{[K]}/1$ en régime stationnaire, alors le temps moyen d'attente dans la file, $W_q^{(K)}$, exprimé en unités de temps normalisées ($\lambda = 1$), est donné par

$$W_q^{(K)} = \sum_{i=1}^{K-1} (1 - z_i)^{-1} + \frac{1 - K(1-\rho)}{2(1-\rho)} \qquad (15)$$

où les z_i ($i = 1, \ldots, K-1$) sont les (K-1) racines à l'intérieur (strictement) du disque-unité de l'équation caractéristique

$$z^K e^{K\rho(1-z)} - 1 = 0 \ . \qquad (16)$$

5.7.3 Lien avec les files d'attente $E_k/E_k/1$

Downton (1956) avait souligné le fait que l'équation caractéristique associée à une quasi-file d'attente $Q\text{-}M/D^{[K]}/1$ est la même que celle associée à une file d'attente classique $E_K/D/1$ ayant même taux d'arrivée et même coefficient d'achalandage; une comparaison rapide des équations caractéristiques associées aux quasi-files $Q\text{-}M/E_k^{[K]}/1$ et aux files $E_K/E_k/1$ nous permettra de voir que cette identité tient aussi pour ces systèmes.

Pour débuter, nous allons énoncer deux théorèmes relatifs au temps moyen passé dans le système pour les files d'attente $E_K/E_k/1$ et $E_K/D/1$ en régime stationnaire. Ces deux théorèmes découlent des résultats de Kotiah, Thompson et Waugh (1969).

Théorème 5.5

Soit ρ (<1) le coefficient d'achalandage d'une file d'attente $E_K/E_k/1$ en régime stationnaire, alors le temps moyen passé dans le système, W', exprimé en unités de temps normalisées ($\lambda = 1$), est donné par :

$$W' = - \sum_{j=K}^{K+k-1} \frac{1}{s_j} \qquad (17)$$

où les s_j ($j = K, \ldots, K+k-1$) sont les k racines avec partie réelle négative de l'équation caractéristique

$$(1 - s/K)^K (1 + \rho s/k)^k - 1 = 0 \qquad (18)$$

Théorème 5.6

Soit ρ (<1) le coefficient d'achalandage d'une file d'attente $E_K/D/1$ en régime stationnaire, alors, le temps moyen passé dans le système, W', exprimé en unités de temps normalisées ($\lambda = 1$), est donnée par :

$$W' = \sum_{i=1}^{K-1} \frac{1}{s_i} + \frac{1}{2(1-\rho)} [1/K - (1-\rho)^2] \qquad (19)$$

où les $s_i (i = 1, \ldots, K-1)$ sont les $(K-1)$ racines avec partie réelle positive de l'équation caractéristique

$$(1 - s/K)^K e^{s\rho} - 1 = 0. \qquad (20)$$

Nous avons maintenant en main tous les éléments nécessaires pour la démonstration de ce que nous avancions au début de cette sous-section.

Théorème 5.7

Soit $W_q^{(K)}$ le temps moyen d'attente dans la file d'une quasi-file d'attente $Q\text{-}M/E_k^{[K]}/1$ (resp. $Q\text{-}M/D^{[K]}/1$) en régime stationnaire et soit W' le temps moyen passé dans une file d'attente $E_K/E_k/1$ (resp. $E_K/D/1$) en régime stationnaire, si les deux systèmes ont taux d'arrivée unitaire et même coefficient de remplissage ρ, alors

$$\frac{W_q^{(K)}}{K} = W' - \rho \left(\frac{k-1}{2k}\right) \qquad (21) \quad (\text{cas } Q\text{-}M/E_k^{[K]}/1)$$

ou

$$\frac{W_q^{(K)}}{K} = W' - \frac{\rho}{2} \qquad (22) \quad (\text{cas } Q\text{-}M/D^{[K]}/1) \ .$$

Démonstration

Nous allons démontrer le théorème séparément pour chacune des deux classes de quasi-files d'attente.

Cas $Q\text{-}M/E_k^{[K]}/1$

Nous devons d'abord démontrer que les équations caractéristiques définies dans les théorèmes 5.3 et 5.5 sont équivalentes. Pour ce faire, nous effectuons le changement de variable $z = 1 - s/K$ (c'est-à-dire $s = K(1-z)$) dans l'équation (18) et nous retrouvons (14).

Il s'agit maintenant de montrer que les k racines avec partie réelle négative de (18) correspondent bien aux k racines à l'extérieur du disque-unité de (14). Considérons une racine s_j ($j \in \{K,\ldots,K+k-1\}$) quelconque de (18). Tout d'abord, il est bien clair que si s_j est une racine de (18), $(1 - s_j/K)$ sera une racine de (14). D'autre part, comme $Re(s_j) < 0$, il est évident que $Re(1 - s_j/K) > 1$, ce qui implique que $|1 - s_j/K| > 1$ et donc que $(1 - s_j/K)$ se trouve à l'extérieur du disque-unité. Il suffit maintenant de remarquer que le changement de variable effectué est une bijection de C dans lui-même et que les deux ensembles de racines qui nous intéressent sont de cardinalité k, pour conclure à la correspondance de $\{s_j\}_{j=K}^{K+k-1}$ et de $\{z_j\}_{j=k}^{K+k-1}$. Pour la suite, nous posons, sans perte de généralité, $z_j = 1 - s_k/K$ ($j = K, \ldots, K+k-1$).

On sait par le théorème 5.5 que

$$W' = \sum_{j=K}^{K+k-1} \frac{1}{s_j} \quad .$$

Si l'on effectue la substitution $s_j = K(1-z_j)$, on obtient

$$W' = -\frac{1}{K} \sum_{j=K}^{K+k-1} (1-z_k)^{-1}$$

$$\Rightarrow \quad K\,W' = \sum_{j=K}^{K+k-1} (z_k-1)^{-1} \qquad (23)$$

Par le théorème 5.3, on a

$$W_q^{(K)} = \sum_{j=K}^{K+k-1} (z_j-1)^{-1} - K\rho\,(\frac{k-1}{2k}) \quad . \qquad (13)$$

Substituant (23) dans (13), on obtient

$$W_q^{(K)} = K\,W' - K\rho\,(\frac{k-1}{2k})$$

$$\Rightarrow \quad \frac{W_q^{(K)}}{K} = W' - \rho\,(\frac{k-1}{2k})$$

Ceci complète la preuve pour le cas Q-M/$E_k^{[K]}$/1.

Cas Q-M/$D^{[K]}$/1

En utilisant ici aussi le changement de variable $z = 1 - s/K$, on montre facilement l'équivalence des équations caractéristiques (16) et (20).

Pour montrer l'équivalence entre les deux ensembles de racines, il faut procéder dans le sens inverse (c'est-à-dire à partir des z_i pour aller vers les s_i) de celui que nous avons utilisé auparavant. En effet, il est facile de voir que $|z_i| < 1 \Rightarrow Re(z_i) < 1 \Rightarrow Re(K(1-z_i)) > 0$. A part cette petite différence, la correspondance entre les deux ensembles de racines est établie comme dans le cas précédent. On peut donc poser

$z_i = 1 - s_i/K \quad (i=1,\ldots,K-1)$,

Par le théorème 5.6, on a

$$W' = \sum_{i=1}^{K-1} \frac{1}{s_i} + \frac{1}{2(1-\rho)} [\, 1/K-(1-\rho)^2 \,]$$

$$\Rightarrow \quad W' = \frac{1}{K} \sum_{i=1}^{K-1} (1-z_i)^{-1} + \frac{1}{2(1-\rho)} [\, 1/K - (1-\rho)^2 \,]$$

$$\Rightarrow \quad KW' = \sum_{i=1}^{K-1} (1-z_i)^{-1} + \frac{1}{2(1-\rho)} [\, 1 - K(1-\rho)^2 \,] \; . \qquad (23)$$

D'autre part, on sait par le théorème 5.4 que

$$W_q^{(k)} = \sum_{i=1}^{K-1} (1-z_i)^{-1} \quad \frac{1-K(1-\rho)}{2(1-\rho)} \; . \qquad (15)$$

Soustrayant membre à membre (23) de (15), on obtient

$$W_q^{(k)} - KW' = \frac{1}{2(1-\rho)} \{ [\, 1-K(1-\rho)] \; - [\, 1-K(1-\rho)^2]$$

$$W_q^{(k)} - KW' = - \frac{K\rho}{2}$$

$$\Leftrightarrow \quad \frac{W_q^{(k)}}{K} = W' - \frac{\rho}{2} \; .$$

Ceci complète la preuve pour le cas Q-M/$D^{[K]}$/1.

□

Corollaire 5.2

Soit $W_q^{(K)}$ le temps moyen d'attente dans la file d'une quasi-file d'attente Q-M/$E_k^{[K]}$/1 (resp. Q-M/$D^{[K]}$/1) en régime stationnaire et soit W_q' le temps moyen d'attente dans une file d'attente $E_K/E_k/1$ (resp. $E_K/D/1$) en régime stationnaire, si les deux systèmes ont taux d'arrivée unitaire et même coefficient de remplissage ρ, alors

$$\frac{W_q^{(K)}}{K} = W_q' + \rho\left(\frac{k+1}{2k}\right) \quad (24) \quad \text{(cas Q-M/}E_k^{[K]}\text{/1)}$$

$$\text{ou} \quad \frac{W_q^{(K)}}{K} - = W_q' + \frac{\rho}{2} \quad (25) \quad \text{(cas Q-M/}D^{[K]}\text{/1)}$$

Démonstration

Dans une file d'attente avec service individuel, si on a un taux d'arrivée unitaire, le temps moyen passé dans le système (W') et le temps moyen d'attente satisfont la relation $W' = W_q' + \rho$. Il suffit alors de substituer le terme $(W_q' + \rho)$ à W' dans les relations (21) et (22) du théorème 5.7 pour démontrer le corollaire.

□

On peut remarquer dans le théorème 5.7 et dans le corollaire 5.2 que les expressions pour $W_q^{(K)}$ dans une quasi-file Q-M/$D^{[K]}$/1 pourraient être interprétées comme les limites quand $k \to \infty$ des expressions que l'on obtient pour les quasi-files Q-M/$E_k^{[K]}$/1.

Si nous avions une formule exacte pour le temps d'attente dans les files $E_K/E_k/1$ et $E_K/D/1$, tous nos problèmes seraient réglés; malheureusement, tel n'est pas le cas.

Il existe néanmoins deux manières de tirer profit des résultats du corollaire 5.2 :

a) en utilisant les tables qui existent pour le temps d'attente dans les files $E_K/E_k/1$ et $E_K/D/1$ (entre autres, les tables de Hillier et Lo, 1971, et celles de Hillier et Yu, 1981) pour calculer $W_q^{(K)}$;

b) en construisant des approximations pour $W_q^{(K)}$ à partir d'approximations pour W_q'.

5.7.4 Approximation de base

Il est bien connu (voir, entre autres, Kleinrock, 1975, p. 306) que le temps d'attente dans une file G/G/1 en régime stationnaire ($\rho < 1$) peut s'écrire sous la forme

$$W_q' = \frac{\sigma_A^2 + \sigma_S^2 + (1/\lambda)^2(1-\rho)^2}{2(1/\lambda)(1-\rho)} - \frac{E[I^2]}{2E[I]} \qquad (26)$$

où σ_A^2 est la variance de la distribution des intervalles inter-arrivées,

σ_S^2 est la variance de la distribution des intervalles de service,

λ est le taux d'arrivée

et I est la variable aléatoire "longueur d'une période creuse pour le système étudié".

Pour une file d'attente $E_K/E_k/1$ avec taux d'arrivée unitaire, on a

$$\sigma_A^2 = \frac{1}{K} ,$$

$$\sigma_S^2 = \frac{\rho^2}{k}$$

et $\lambda = 1$,

ce qui fait que l'on peut réécrire (26) sous la forme

$$W_q' = \frac{(1/K) + (\rho^2/k) + (1-\rho)^2}{2(1-\rho)} - \frac{E[I^2]}{2E[I]} \qquad (27)$$

Pour une file d'attente $E_K/D/1$ avec taux d'arrivée unitaire, (26) devient

$$W_q' = \frac{(1/K) + (1-\rho)^2}{2(1-\rho)} - \frac{E[I^2]}{2E[I]} \qquad (28)$$

Revenons maintenant pour quelques instants à nos quasi-files d'attente. Si, dans la relation (24), on remplace W_q' par sa valeur telle que donnée par (27), on obtient pour les quasi-files $Q\text{-}M/E_k^{[K]}/1$,

$$W_q^{(K)} = \frac{(1/K) + (\rho^2/k) + (1-\rho)^2}{2(1-\rho)} - \frac{E[I^2]}{2E[I]} + \rho\left(\frac{k+1}{2k}\right)$$

$$\Rightarrow \qquad \frac{W_q^{(K)}}{K} = \frac{1}{2} + \frac{(1/K) + (\rho/k)}{2(1-\rho)} - \frac{E[I^2]}{2E[I]} \qquad (29)$$

Pour les quasi-files $Q\text{-}M/D^{[K]}/1$, on tire de (25) et (28)

$$\frac{W_q^{(K)}}{K} = \frac{(1/K) + (1-\rho)^2}{2(1-\rho)} - \frac{E[I^2]}{2E[I]} + \frac{\rho}{2}$$

$$\Rightarrow \qquad \frac{W_q^{(K)}}{K} = \frac{1}{2} + \frac{1}{2K(1-\rho)} - \frac{E[I^2]}{2E[I]} \qquad (30)$$

Les relations (29) et (30) sont intéressantes, mais nous ne connaissons pas, en général, la valeur des termes $\frac{E[I^2]}{2E[I]}$. Si nous étions capables de trouver une valeur approximative pour ces termes, cela nous donnerait une approximation de $W_q^{(K)}$. Il va donc falloir nous pencher sur la distribution de la longueur des périodes creuses dans les files $E_K/E_k/1$ et $E_K/D/1$. Si l'on observe une file d'attente G/G/1 quelconque, on peut remarquer les faits suivants :

a) Chaque période creuse est contenue strictement dans un intervalle interarrivées donné.

b) Pour qu'un intervalle interarrivées engendre une période creuse, il faut que le serveur réussisse à "vider" le système de tous les usagers qui s'y trouvaient immédiatement après le début de l'intervalle.

c) Moins il y a de personnes dans le système au début d'un intervalle interarrivées, plus il y a de chances que cet intervalle engendre une période creuse.

Si l'on pousse à sa limite le raisonnement sous-tendant la remarque c), on s'aperçoit que la situation dans laquelle on a le plus de chances qu'un intervalle interarrivées engendre une période creuse, est celle où il n'y a qu'un seul client (celui qui vient tout juste d'arriver), c'est-à-dire quand le début de l'intervalle de service et le début de l'intervalle interarrivées coïncident et que la file est vide. Ceci nous amène à la conjecture suivante :

Conjecture 5.1

La majorité des périodes creuses d'une file G/G/1 en régime stationnaire sont engendrées par l'arrivée d'un client alors que le système était vide au préalable.

Evidemment, la formulation de cette conjecture est vague (qu'entendons-nous exactement par "majorité"?), mais, à ce stade-ci de nos recherches, il nous paraît difficile d'en arriver à un énoncé plus précis.

Si presque toutes les périodes creuses sont engendrées par l'arrivée d'un client dans un système vide, on pourra se servir de la distribution de ces périodes creuses pour approximer celle de I. Nous allons donc étudier la distribution de la variable aléatoire Z, "longueur d'une période creuse engendrée par l'arrivée d'un client alors que le système est vide". Si on appelle X la V.A. "longueur d'un intervalle interarrivées", Y la V.A. "longueur d'un intervalle de service" et $F_Z(t)$, $F_X(t)$ et $F_Y(t)$ les distribution cumulatives de Z, X et Y, on a

$$\Pr\left\{\begin{matrix}\text{intervalle observé engendre}\\ \text{une période creuse}\end{matrix}\right\} = \Pr\{X > Y\}$$

$$\Pr\left\{\begin{matrix}\text{intervalle observé engendre}\\ \text{une période creuse de}\\ \text{longueur} > t\end{matrix}\right\} = \Pr\{X > Y + t\},\quad t \geq 0$$

$$\Rightarrow \quad \Pr\{Z > t\} = \frac{\Pr\{X > Y + t\}}{\Pr\{X > Y\}}, \qquad t \geq 0$$

$$\Leftrightarrow \quad 1 - F_Z(t) = \frac{\Pr\{X - Y > t\}}{\Pr\{X - Y > 0\}}, \qquad t \geq 0$$

(Remarque : comme $\rho < 1$, on a $E[X] > E[Y] \Rightarrow \Pr\{X > Y\} > 0$)

On veut se servir de la distribution de Z pour "remplacer" celle de I de manière à approximer dans (29) et (30) la terme $\frac{E[I^2]}{2E[I]}$ par $\frac{E[Z^2]}{2E[Z]}$.

Z étant, par définition, une V.A. non-négative, on peut calculer ses moments par la formule

$$E[Z^r] = r \int_0^\infty t^{r-1} [1-F_Z(t)]\, dt$$

(Feller, 1966, p. 150).

Il s'ensuit que

$$\frac{E[Z^2]}{2E[Z]} = \frac{2\int_0^\infty t\,[1-F_Z(t)]\,dt}{2\int_0^\infty [1-F_Z(t)]\,dt}$$

$$= \frac{\int_0^\infty t\ \frac{Pr\{X-Y>t\}}{Pr\{X-Y>0\}}\ dt}{\int_0^\infty \frac{Pr\{X-Y>t\}}{Pr\{X-Y>0\}}\ dt}$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t\ Pr\{X-Y>t\}\ dt}{\int_0^\infty Pr\{X-Y>t\}\ dt} \qquad (31)$$

X et Y étant des variables aléatoires indépendantes, on obtient en conditionnant sur la valeur de Y,

$$Pr\{X-Y>t\} = \int_0^\infty Pr\{X>t+y \mid Y=y\}\, dF_Y(y)$$

$$= \int_0^\infty Pr\{X>t+y\}\, dF_Y(y)$$

$$Pr\{X-Y>t\} = \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)$$

Substituant dans (31), on trouve

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)\, dt}{\int_0^\infty \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)\, dt} \qquad (32)$$

Pour pouvoir poursuivre notre analyse, il faut spécifier les distributions de X et Y. Pour la file $E_K/E_k/1$ à taux d'arrivée unitaire et à coefficient d'achalandage ρ, X suit une loi d'Erlang de moyenne 1 et de paramètre de forme K et Y suit une loi d'Erlang de moyenne ρ et de paramètre de forme k, ce qui implique que

$$F_X(x) = 1 - e^{-Kx} \sum_{i=0}^{K-1} \frac{(Kx)^i}{i!} \quad , \quad x \geq 0,$$

et $$\frac{dF_Y(y)}{dy} = \frac{(k/\rho)^k}{(k-1)!}\, y^{k-1}\, e^{-(k/\rho)y}, \quad y \geq 0.$$

Calculons dans un premier temps,

$$G^* = \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)$$

$$= \int_0^\infty [e^{-K(t+y)} \sum_{i=0}^{K-1} \frac{[K(t+y)]^i}{i!}] \frac{(k/\rho)^k}{(k-1)!}\, y^{k-1}\, e^{-(k/\rho)y}\, dy$$

$$G^* = e^{-Kt} (\tfrac{k}{\rho})^k \sum_{i=0}^{K-1} K^i \int_0^\infty \frac{(t+y)^i}{i!} \frac{y^{k-1}}{(k-1)!}\, e^{-[K+k/\rho]y}\, dy$$

Posons $G_{i\ell} = \int_0^\infty \frac{(t+y)^i}{i!} \frac{y^\ell}{\ell!}\, e^{-[K+k/\rho]}\, dy$. On a alors

$$G^* = e^{-Kt} (\tfrac{k}{\rho})^k \sum_{i=0}^{K-1} K^i\, G_{i(k-1)} \qquad (33)$$

Pour $i \geq 1$ et $\ell \geq 1$, on obtient, en intégrant par parties,

$$G_{i\ell} = \left[\frac{(t+y)^i}{i!} \frac{y^\ell}{\ell!} \frac{e^{-[K+k/\rho]y}}{-(K+k/\rho)} \right]_{y=0}^{\infty}$$

$$- \int_0^\infty \frac{(t+y)^{i-1}}{(i-1)!} \frac{y^\ell}{\ell!} \frac{e^{-[K+k/\rho]}}{-(K+k/\rho)} \, dy$$

$$- \int_0^\infty \frac{(t+y)^i}{i!} \frac{y^{\ell-1}}{(\ell-1)!} \frac{e^{-[K+k/\rho]}}{-(K+k/\rho)} \, dy$$

Pour $\ell \geq 1$, le premier terme est nul et donc

$$G_{i\ell} = \frac{1}{K+k/\rho} \left(G_{(i-1)\ell} + G_{i(\ell-1)} \right) \tag{34}$$

Pour $i \geq 1$ et $\ell = 0$,

$$G_{i0} = \left[\frac{(t+y)^i}{i!} \frac{e^{-(K+k/\rho)y}}{-(K+k/\rho)} \right]_{y=0}^{\infty} - \int_0^\infty \frac{(t+y)^{i-1}}{(i-1)!} \frac{e^{-(K+k/\rho)y}}{-(K+k/\rho)} \, dy$$

$$\Rightarrow \quad G_{i0} = \frac{1}{K+k/\rho} \left(\frac{t^i}{i!} + G_{(i-1)0} \right) \tag{35}$$

Pour $i = 0$ et $\ell \geq 1$,

$$G_{0\ell} = \left[\frac{y^\ell}{\ell!} \frac{e^{-(K+k/\rho)y}}{-(K+k/\rho)} \right]_{y=0}^{\infty} - \int_0^\infty \frac{y^{\ell-1}}{(\ell-1)!} \frac{e^{-(K+k/\rho)}}{-(K+k/\rho)} \, dy$$

$$\Rightarrow \quad G_{0\ell} = \frac{1}{K+k/\rho} \tag{36}$$

Pour $i = \ell = 0$

$$G_{00} = \left[\frac{e^{-(K+k/\rho)y}}{-(K+k/\rho)}\right]_{y=0}^{\infty} = \frac{1}{K+k/\rho} \tag{37}$$

On pourrait essayer de calculer directement $G_{i\ell}$ pour n'importe quelle valeur de i et de ℓ en utilisant les relations de récurrence (34) à (37), mais il y a plus simple. Posons plutôt

$$\tilde{G}_{i\ell} = (K+k/\rho)^{(i+\ell+1)}\, G_{i\ell}$$

et réécrivons les relations (34) à (37) en termes des $\tilde{G}_{i\ell}$. On obtient

$$\tilde{G}_{i\ell} = \tilde{G}_{(i-1)\ell} + \tilde{G}_{i(\ell-1)}, \qquad i \geq 1,\ \ell \geq 1 \tag{38}$$

$$\tilde{G}_{i0} = (K+k/\rho)^{i}\,\frac{t^{i}}{i!} + \tilde{G}_{(i-1)0}, \qquad i \geq 1, \tag{39}$$

$$\tilde{G}_{0\ell} = \tilde{G}_{0(\ell-1)}, \qquad \ell \geq 1, \tag{40}$$

et $$G_{00} = 1. \tag{41}$$

On en tire d'abord les résultats

$$\tilde{G}_{i0} = \sum_{j=0}^{i} \frac{[(K+k/\rho)t]^{j}}{j!}, \qquad i \geq 1, \tag{42}$$

et $$\tilde{G}_{0\ell} = 1, \qquad \ell \geq 0. \tag{43}$$

Soit $\bar{G}_{i\ell} = \sum_{j=0}^{i} \binom{(i-j)+\ell}{\ell} \frac{[(K+k/\rho)t]^{j}}{j!}, \qquad i \geq 0,\ \ell \geq 0.$

Nous allons montrer que $\tilde{G}_{i\ell} = \bar{G}_{i\ell}$, $i \geq 0,\ \ell \geq 0$.

Pour $i \geq 1$,

$$\bar{G}_{i0} = \sum_{j=0}^{i} \binom{(i-j)+0}{0} \frac{[(K+k/\rho)t]^j}{j!} = \sum_{j=0}^{i} \frac{[(K+k/\rho)t]^j}{j!} = \tilde{G}_{i0}$$

Pour $\ell \geq 0$,

$$\bar{G}_{0\ell} = \binom{\ell}{\ell} \frac{[(K+k/\rho)t]^0}{0!} = 1 = \tilde{G}_{0\ell}.$$

Pour $i \geq 1$, $\ell \geq 1$,

$$\bar{G}_{i\ell} = \sum_{j=0}^{i} \left[\binom{(i-1)-j+\ell}{\ell} + \binom{(i-j)+(\ell-1)}{\ell-1} \right] \frac{[(K+k/\rho)t]^j}{j!}$$

$$= \sum_{j=0}^{i} \binom{(i-1)-j+\ell}{\ell} \frac{[(K+k/\rho)t]^j}{j!} + \sum_{j=0}^{i} \binom{(i-j)+(\ell-1)}{\ell-1} \frac{[(K+k/\rho)t]^j}{j!}$$

$$= \binom{\ell-1}{\ell} \frac{[(K+k/\rho)t]^i}{i!} + \bar{G}_{(i-1)\ell} + \bar{G}_{i(\ell-1)}$$

$$\Rightarrow \quad \bar{G}_{i\ell} = \bar{G}_{(i-1)\ell} + \bar{G}_{i(\ell-1)} \quad , \quad \text{car} \quad \binom{\ell-1}{\ell} = 0 \ .$$

On voit donc que les $\bar{G}_{i\ell}$, $i \geq 0$, $\ell \geq 0$, satisfont les relations (38), (42) et (43), ce qui nous permet de conclure que $\bar{G}_{i\ell} = \tilde{G}_{i\ell}$, $i \geq 0$, $\ell \geq 0$.

Maintenant que l'on a obtenu une expression pour $\tilde{G}_{i\ell}$, on peut en déduire

$$G_{i\ell} = (K+k/\rho)^{-(i+\ell+1)} \sum_{j=0}^{i} \binom{(i-j)+\ell}{\ell} \frac{[(K+k/\rho)t]^j}{j!} \ , \quad i \geq 0, \ \ell \geq 0.$$

Substituant dans (33), on obtient

$$G^* = e^{-Kt} \left(\frac{k}{\rho}\right)^k \sum_{i=0}^{K-1} K^i \, (K+k/\rho)^{-(i+k)} \sum_{j=0}^{i} \binom{(i-j)+(k-1)}{k-1} \frac{[(K+k/\rho)t]^j}{j!}$$

$$= e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{i=0}^{K-1} \sum_{j=0}^{i} K^{i-j} K^j (K+k/\rho)^{-i} \binom{(i-j)+(k-1)}{k-1} \frac{[(K+k/\rho)t]^j}{j!}$$

$$= e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{i=0}^{K-1} \sum_{j=0}^{i} \frac{(Kt)^j}{j!} \binom{(i-j)+(k-1)}{k-1} K^{i-j} (K+k/\rho)^{-(i-j)}$$

$$= e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \sum_{i=j}^{K-1} \binom{(i-j)+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i-j}$$

$$G^* = e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \sum_{i=0}^{K-1-j} \binom{i'+(k-1)}{i'} \left(\frac{K\rho}{K\rho+k}\right)^{i'} \tag{44}$$

Posons $\xi_j = \sum_{i'=0}^{K-1-j} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'}$.

Alors $G^* = e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \xi_j$. (45)

Connaissant G^*, nous pouvons calculer $\frac{E[Z^2]}{2E[Z]}$.

On a

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t\, G^*\, dt}{\int_0^\infty G^*\, dt}$$

$$= \frac{\int_0^\infty t\{e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{j=0}^{K-1} \frac{(Kt)^t}{j!} \xi_j\}\, dt}{\int_0^\infty \{e^{-Kt} \left(\frac{k}{K\rho+k}\right)^k \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \xi_j\}\, dt}$$

$$= \frac{\int_0^\infty t\, e^{-Kt} \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \xi_j\, dt}{\int_0^\infty e^{-Kt} \sum_{j=0}^{K-1} \frac{(Kt)^j}{j!} \xi_j\, dt}$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{j=0}^{K-1} \xi_j \int_0^\infty t \frac{(Kt)^j}{j!} e^{-Kt}\, dt}{\sum_{j=0}^{K-1} \xi_j \int_0^\infty \frac{(Kt)^j}{j!} e^{-Kt}\, dt} \tag{46}$$

On a $\int_0^\infty t \frac{(Kt)^j}{j!} e^{-Kt}\, dt = \frac{K^j}{j!} \int_0^\infty t^{j+1} e^{-Kt}\, dt = \frac{K^j}{j!} \frac{(j+1)}{K^{j+2}} = \frac{j+1}{K^2}$

et $\int_0^\infty \frac{(Kt)^j}{j!} e^{-Kt}\, dt = \frac{K^j}{j!} \int_0^\infty t^j e^{-Kt}\, dt = \frac{K^j}{j!} \frac{j!}{K^{j+1}} = \frac{1}{K}$.

On en tire

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{j=0}^{K-1} \xi_j \frac{(j+1)}{K^2}}{\sum_{j=0}^{K-1} \xi_j \left(\frac{1}{K}\right)} = \frac{\sum_{j=0}^{K-1} (j+1)\, \xi_j}{K \sum_{j=0}^{K-1} \xi_j} \tag{47}$$

Développons séparément le numérateur et le dénominateur de (47).

Pour le numérateur on a

$$\sum_{j=0}^{K-1} (j+1)\, \xi_j = \sum_{j=0}^{K-1} (j+1) \sum_{i=0}^{K-1-j} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'}$$

$$= \sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} \sum_{j=0}^{K-1-i'} (j+1)$$

$$\sum_{j=0}^{K-1} (j+1)\, \xi_j = \sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} \frac{(K-i')(K-i'+1)}{2}$$

Pour le dénominateur, on a

$$K \sum_{j=0}^{K-1} \varepsilon_j = K \sum_{j=0}^{K-1} \sum_{i'=0}^{K-1-j} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'}$$

$$= K \sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} \sum_{j=0}^{K-1-i'} 1$$

$$K \sum_{j=0}^{K-1} \varepsilon_j = K \sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')$$

On peut donc réécrire $\frac{E[Z^2]}{2E[Z]}$ sous la forme

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} \frac{(K-i')(K-i'+1)}{2}}{K \sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')}$$

ou encore

$$\frac{E[Z^2]}{2E[Z]} = \frac{1}{2K} \left\{ 1 + \frac{\sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')^2}{\sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')} \right\} \qquad (48)$$

Malgré l'aspect rébarbatif de (48), il est relativement simple de calculer $\frac{E[Z^2]}{2E[Z]}$. En effet, si l'on pose $a_{i'} = \binom{i'+(k-1)}{k-1}\left(\frac{K\rho}{K\rho+k}\right)^{i'}$, $i'=0,\ldots,K-1$, on a les relations suivantes :

$$\cdot\ a_0 = \binom{k-1}{k-1}\left(\frac{K\rho}{K\rho+k}\right)^0 = 1$$

$$\cdot\ a_{i'+1} = \binom{(i'+1)+(k-1)}{k-1}\left(\frac{K\rho}{K\rho+k}\right)^{i'+1}$$

$$= \frac{[(i'+1)+(k-1)]!}{(k-1)!\,(i'+1)!} \left(\frac{K\rho}{K\rho+k}\right)^{i'+1}$$

$$= \left(\frac{K\rho}{K\rho+k}\right) \frac{[(i'+1)+(k-1)]}{(i'+1)} \frac{[i'+(k-1)]}{(k-1)!\,i'!} \left(\frac{K\rho}{K\rho+k}\right)^{i'}$$

$$a_{i'+1} = \left(\frac{K\rho}{K\rho+k}\right) \frac{[(i'+1)+(k-1)]}{(i'+1)} a_{i'} \quad , \; i' \geq 0$$

$$\cdot \frac{E[Z^2]}{2E[Z]} = \frac{1}{2K} \left[1 + \frac{\sum_{i'=0}^{K-1} a_{i'} (K-i')^2}{\sum_{i'=0}^{K-1} a_{i'} (K-i')}\right] \tag{49}$$

Il suffit donc d'une simple calculatrice programmable pour calculer $\frac{E[Z^2]}{2E[Z]}$, même pour des valeurs assez importantes du paramètre K. Etant donné que, pour chaque valeur de l'indice i', le nombre d'opérations à effectuer est constant, on peut conclure que le temps de calcul requis pour l'évaluation de $\frac{E[Z^2]}{2E[Z]}$ croît linéairement avec K.

Nous indiquerons un peu plus loin dans cette sous-section la forme exacte de l'approximation obtenue pour $W_q^{(K)}$ quand on remplace $\frac{E[I^2]}{2E[I]}$ par $\frac{E[Z^2]}{2E[Z]}$.

Quand le paramètre de forme de la distribution des intervalles interarrivées tend vers l'infini, on obtient à la limite une distribution déterministe. Il parait donc important d'évaluer $\frac{E[Z^2]}{2E[Z]}$ pour une file $D/E_k/1$.

On se rappelle que la forme générale de $\frac{E[Z^2]}{2E[Z]}$ est

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)\, dt}{\int_0^\infty \int_0^\infty [1-F_X(t+y)]\, dF_Y(y)\, dt}$$

On a toujours

$$\frac{dF_Y(y)}{dy} = \frac{(k/\rho)^k}{(k-1)!}\, y^{k-1}\, e^{-(k/\rho)y}\ ,\quad y \geq 0.$$

mais maintenant

$$F_X(x) = \begin{cases} 0\ , & x < 1\ , \\ 1\ , & x \geq 1\ . \end{cases}$$

C'est donc dire que pour $t \geq 1$ et y quelconque, on a toujours $[1-F_X(t+y)] = 0$. De même, pour $t \in [0,1)$, on a $[1-F_X(t+y)] = 0$ pour $y \geq 1-t$ et $[1-F_X(t+y)] = 1$ pour $y \in [0,1-t)$. Il s'ensuit alors que

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^{1^-} t \int_0^{(1-t)^-} dF_Y(y)dt}{\int_0^{1^-} \int_0^{(1-t)^-} dF_Y(y)}$$

Etant donnée la distribution de Y, on peut inclure les limites supérieures d'intégration sans changer la valeur des intégrales calculées. On a donc

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^1 t \int_0^{1-t} dF_Y(y)}{\int_0^1 \int_0^{1-t} dF_Y(y)}$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^1 t\, F_Y(1-t)\, dt}{\int_0^1 F_Y(1-t)\, dt}$$

(Y, V.A. non négative => $F_Y(0)=0$)

On peut maintenant effectuer le changement de variable $t' = 1 - t$ et inverser l'ordre d'intégration. On obtient alors

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^1 (1-t')\, F_Y(t')\, dt'}{\int_0^1 F_Y(t')\, dt'}$$

Pour alléger la notation, nous laissons tomber le "'" sur "t" dans la suite.

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^1 (1-t)\, F_Y(t)\, dt}{\int_0^1 F_Y(t)\, dt} = 1 - \frac{\int_0^1 t\, F_Y(t) dt}{\int_0^1 F_Y(t)\, dt}$$

Y étant une V.A. suivant une loi d'Erlang de moyenne ρ et de paramètre k, on a

$$F_Y(y) = 1 - e^{-ky/\rho} \sum_{i=0}^{k-1} \frac{(ky/\rho)^i}{i!}, \qquad y \geq 0 .$$

Et alors,

$$\frac{E[Z^2]}{2E[Z]} = 1 - \frac{\int_0^1 t \left[1 - e^{-kt/\rho} \sum_{i=0}^{k-1} \frac{(kt/\rho)^i}{i!}\right] dt}{\int_0^1 \left[1 - e^{-kt/\rho} \sum_{i=0}^{k-1} \frac{(kt/\rho)^i}{i!}\right] dt}$$

$$= 1 - \frac{\int_0^1 t\,dt - \int_0^1 te^{-kt/\rho} \sum_{i=0}^{k-1} \frac{(kt/\rho)^i}{i!}\,dt}{\int_0^1 dt - \int_0^1 e^{-kt/\rho} \sum_{i=0}^{k-1} \frac{(kt/\rho)^i}{i!}\,dt}$$

$$= 1 - \frac{(1/2) - \sum_{i=0}^{k-1} \int_0^1 t\, e^{-kt/\rho} \frac{(kt/\rho)^i}{i!}\,dt}{1 - \sum_{i=0}^{k-1} \int_0^1 e^{-kt/\rho} \frac{(kt/\rho)^i}{i!}\,dt}$$

$$\frac{E[Z^2]}{2E[Z]} = 1 - \frac{(1/2) - \sum_{i=0}^{k-1} \frac{\rho}{k}(i+1) \int_0^1 e^{-kt/\rho} \frac{(kt/\rho)^{i+1}}{i!}\,dt}{1 - \sum_{i=0}^{k-1} \int_0^1 e^{-kt/\rho} \frac{(kt/\rho)^i}{i!}\,dt}$$

Si l'on pose $G_i = \int_0^1 e^{-kt/\rho} \frac{(kt/\rho)^i}{i!}\,dt$, on obtient

$$\frac{E[Z^2]}{2E[Z]} = 1 - \frac{(1/2) - \sum_{i=0}^{k-1} \frac{\rho}{k}(i+1)\, G_{i+1}}{1 - \sum_{i=0}^{k-1} G_i} \qquad (50)$$

Pour $i \geq 1$, on a

$$G_i = \left[\frac{e^{-kt/\rho}}{-k/} \frac{(kt/\rho)^i}{i!}\right]_{t=0}^{1} - \int_0^1 \frac{e^{-kt/\rho}}{-k/\rho} \frac{k}{\rho} \frac{(kt/\rho)^{i-1}}{(i-1)!}\,dt$$

$$G_i = -\frac{e^{-k/\rho}(k/\rho)^{i-1}}{i!} + G_{i-1}$$

Par induction, on obtient

$$G_i = \sum_{j=1}^{i} \frac{(-e^{-k/\rho})(k/\rho)^{j-1}}{j!} + G_0$$

$$G_i = -\left(\frac{\rho}{k}\right) e^{-k/\rho} \sum_{j=1}^{i} \frac{(k/\rho)^j}{j!} + G_0$$

Pour ce qui est de G_0, on a

$$G_0 = \int_0^1 e^{-kt/\rho}\, dt = \left[\frac{e^{-kt/\rho}}{-k/}\right]_{t=0}^{1} = \left(\frac{\rho}{k}\right)(1-e^{-k/\rho})$$

On peut en conclure que

$$G_i = \left(\frac{\rho}{k}\right)\left[1-e^{-k/\rho} \sum_{j=0}^{i} \frac{(k/\rho)^i}{j!}\right] \qquad (51)$$

Au numérateur de (50), la sommation devient

$$\sum_{i=0}^{k-1} \frac{\rho}{k}(i+1)\, G_{i+1} = \sum_{i=0}^{k-1} \frac{\rho}{k}(i+1)\left(\frac{\rho}{k}\right)\left[1-e^{-k/\rho} \sum_{j=0}^{i+1} \frac{(k/\rho)^j}{j!}\right]$$

$$= \left(\frac{\rho}{k}\right)^2 \left[\sum_{i=0}^{k-1}(i+1) - e^{-k/\rho} \sum_{i=0}^{k-1}(i+1) \sum_{j=0}^{i+1} \frac{(k/\rho)^j}{j!}\right]$$

$$\sum_{i=0}^{k-1} \frac{\rho}{k}(i+1)\, G_{i+1} = \left(\frac{\rho}{k}\right)^2 \left[\frac{k(k+1)}{2} - e^{-k/\rho} \sum_{i'=1}^{k} i' \sum_{j=0}^{i'} \frac{(k/\rho)^j}{j!}\right]$$

On a

$$\sum_{i'=0}^{k} i' \sum_{j=0}^{i'} \frac{(k/\rho)^j}{j!} = \sum_{i'=0}^{k} i' \sum_{j=0}^{i'} \frac{(k/\rho)^j}{j!} = \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!} \sum_{i'=j}^{k} i' = \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!} \frac{(k-j+1)(k+j)}{2}$$

On conclut

$$\sum_{i=0}^{k-1} \frac{\rho}{k}(i+1)G_{i+1} = \left(\frac{\rho}{k}\right)^2 \left[\frac{k(k+1)}{2} - e^{-k/\rho} \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!} \frac{(k-j+1)(k+j)}{2}\right] \qquad (52)$$

Au dénominateur de (50), la sommation devient

$$\sum_{i=0}^{k-1} G_i = \sum_{i=0}^{k-1} \left(\frac{\rho}{k}\right) \left[1 - e^{-k/\rho} \sum_{j=0}^{i} \frac{(k/\rho)^j}{j!}\right]$$

$$= \left(\frac{\rho}{k}\right) \sum_{i=0}^{k-1} 1 - \left(\frac{\rho}{k}\right) e^{-k/\rho} \sum_{i=0}^{k-1} \sum_{j=0}^{i} \frac{(k/\rho)^j}{j!}$$

$$= \left(\frac{\rho}{k}\right) k - \left(\frac{\rho}{k}\right) e^{-k/\rho} \sum_{j=0}^{k-1} \frac{(k/\rho)^j}{j!} \sum_{i=1}^{k-1} 1$$

$$\sum_{i=0}^{k-1} G_i = \rho - \left(\frac{\rho}{k}\right) e^{-k/\rho} \sum_{j=0}^{k-1} \frac{(k/\rho)^j}{j!} (k-j) \qquad (53)$$

De (52) et (53), on obtient en remplaçant dans (50) les sommations par leur valeur

$$\frac{E[Z^2]}{2E[Z]} = 1 - \frac{1}{2} \frac{1-\left(\frac{\rho}{k}\right)^2 k(k+1)+\left(\frac{\rho}{k}\right)^2 e^{-k/\rho} \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!}(k-j+1)(k+j)}{1 - \rho + \left(\frac{\rho}{k}\right) e^{-k/\rho} \sum_{j=0}^{k-1} \frac{(k/\rho)^j}{j!}(k-j)} \qquad (54)$$

Par une technique semblable à celle que nous avons proposée dans le cas de la file $E_K/E_k/1$, on peut évaluer $\frac{E[Z^2]}{2E[Z]}$ relativement facilement.

Considérons maintenant le cas où les intervalles de service sont déterministes, notamment celui des files $E_K/D/1$. Quand les intervalles de service suivent une telle loi, toute la masse de la distribution de probabilité est concentrée au point $y = \rho$ et l'on obtient

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t\,[1 - F_X(t+\rho)]\,dt}{\int_0^\infty [1-F_X(t+\rho)]\,dt} .$$

On sait que $F_X(x) = 1 - e^{Kx} \sum_{i=0}^{K-1} \frac{(Kx)^i}{i!}$, $x > 0$. On en tire

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_0^\infty t\, e^{-K(t+\rho)} \sum_{i=0}^{K-1} \frac{[K(t+\rho)]^i}{i!}\,dt}{\int_0^\infty e^{-K(t+\rho)} \sum_{i=0}^{K-1} \frac{K(t+\rho)^i}{i!}\,dt}$$

Effectuons le changement de variable $t' = t + \rho$. On a

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_\rho^\infty (t'-\rho)\, e^{-Kt'} \sum_{i=0}^{K-1} \frac{(Kt')^i}{i!}\,dt'}{\int_\rho^\infty e^{-Kt'} \sum_{i=0}^{K-1} \frac{(Kt')^i}{i!}\,dt'} .$$

Comme auparavant, allégeons la notation en remplaçant "t'" par "t". Alors

$$\frac{E[Z^2]}{2E[Z]} = \frac{\int_\rho^\infty t\, e^{-Kt} \sum_{i=0}^{K-1} \frac{(Kt)^i}{i!}\,dt}{\int_\rho^\infty e^{-Kt} \sum_{i=0}^{K-1} \frac{(Kt)^i}{i!}\,dt} - \rho$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{i=0}^{K-1} \int_\rho^\infty t\, e^{-Kt} \frac{(Kt)^i}{i!}\,dt}{\sum_{i=0}^{K-1} \int_\rho^\infty e^{-Kt} \frac{(Kt)^i}{i!}\,dt} - \rho$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{i=0}^{K-1} \frac{(i+1)}{K} \int_\rho^\infty e^{-Kt} \frac{(Kt)^{i+1}}{(i+1)!}\,dt}{\sum_{i=0}^{K-1} \int_\rho^\infty e^{-Kt} \frac{(Kt)^i}{i!}\,dt} - \rho$$

Si l'on pose $G_i = \int_{\rho}^{\infty} e^{-Kt} \frac{(Kt)^i}{i!}\, dt$, on obtient

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{i=0}^{K-1} \frac{(i+1)}{K} G_{i+1}}{\sum_{i=0}^{K-1} G_i} - \rho \qquad (55)$$

On a pour $i \geq 1$

$$G_i = \left[\frac{e^{-Kt}}{-K} \frac{(Kt)^i}{i!}\right]_{t=\rho}^{\infty} - \int_{\rho}^{\infty} \frac{e^{-Kt}}{-K} K \frac{(Kt)^{i-1}}{(i-1)!}\, dt$$

$$G_i = \frac{e^{-K\rho}}{K} \frac{(K\rho)^i}{i!} + G_{i-1}$$

Par induction, on obtient

$$G_i = \frac{e^{-K\rho}}{K} \sum_{j=1}^{i} \frac{(K\rho)^j}{j!} + G_0 .$$

Pour G_0, on a

$$G_0 = \left[\frac{e^{-Kt}}{-K}\right]_{t=\rho}^{\infty} = \frac{e^{-K\rho}}{K} .$$

C'est donc dire que

$$G_i = \frac{e^{-K\rho}}{K} \sum_{j=0}^{i} \frac{(K\rho)^j}{j!} , \quad i \geq 0 .$$

Après substitution, (55) devient

$$\frac{E[Z^2]}{2E[Z]} = \frac{\sum_{i=0}^{K-1} \frac{(i+1)}{K} \cdot \frac{e^{-K\rho}}{K} \sum_{j=0}^{i+1} \frac{(K\rho)^j}{j!}}{\sum_{i=0}^{K-1} \frac{e^{-K\rho}}{K} \sum_{j=0}^{i} \frac{(K\rho)^j}{j!}} - \rho$$

$$= \frac{\sum_{i=0}^{K-1} (i+1) \sum_{j=0}^{i+1} \frac{(K\rho)^j}{j!}}{K \sum_{i=0}^{K-1} \sum_{j=0}^{i} \frac{(K\rho)^j}{j!}} - \rho$$

$$= \frac{\sum_{j=0}^{K} \frac{(K\rho)^j}{j!} \frac{(K-j+1)(K+j)}{2}}{K \sum_{j=0}^{K-1} \frac{(K\rho)^j}{j!} (K-j)} - \rho$$

$$\frac{E[Z^2]}{2E[Z]} = \frac{1}{2K} \frac{\sum_{j=0}^{K} \frac{(K\rho)^j}{j!} (K-j+1)(K+j)}{\sum_{j=0}^{K-1} \frac{(K\rho)^j}{j!} (K-j)} - \rho \qquad (56)$$

On a donc pour $\frac{E[Z^2]}{2E[Z]}$ une expression qui ressemble à ce que l'on avait obtenu pour les files $E_K/E_k/1$ et $D/E_k/1$. On pourra l'évaluer en utilisant une approche semblable à celle que nous avions proposée pour ces files.

Avant de passer aux expressions pour l'approximation de $W_q^{(K)}$, considérons un instant le cas limite de la file D/D/1. Kleinrock (1975, p. 309) fait remarquer que, dans cette file, chaque intervalle inter-arrivée engendre une période creuse de longueur $(1-\rho)$. On a alors $I = Z$ et

$$\frac{E[I^2]}{2E[I]} = \frac{E[Z^2]}{2E[Z]} = \frac{(1-\rho)^2}{2(1-\rho)} = \frac{(1-\rho)}{2} . \qquad (57)$$

Nous sommes maintenant en mesure de définir les formules d'approximation obtenues en remplaçant dans (29) et (30), $\frac{E[I^2]}{2E[I]}$ par $\frac{E[Z^2]}{2E[Z]}$.

Pour les quasi-files Q-M/$E_k^{[K]}$/1, on a

$$\frac{\hat{W}_q^{(K)}}{K} = \frac{1}{2} + \frac{(1/K)+(\rho/k)}{2(1-\rho)} - \frac{1}{2K} \left\{ 1 + \frac{\sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')^2}{\sum_{i'=0}^{K-1} \binom{i'+(k-1)}{k-1} \left(\frac{K\rho}{K\rho+k}\right)^{i'} (K-i')} \right\} \quad (58)$$

(par (29) et (48)).

$$\text{et } \lim_{K\to\infty} \frac{\hat{W}_q^{(K)}}{K} = \frac{1}{2} + \frac{\rho}{2k(1-\rho)} - \left\{ 1 - \frac{1}{2} \frac{1-(\frac{\rho}{k})^2 k(k+1)+(\frac{\rho}{k})^2 e^{-k/\rho} \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!}(k-j+1)(k+j)}{1-\rho+(\frac{\rho}{k})e^{-k/\rho} \sum_{j=0}^{k-1} \frac{(k/\rho)^j}{j!}(k-j)} \right\} \quad (59$$

(par (29) et (54)).

Pour les quasi-files Q-M/$D^{[K]}$/1, on obtient

$$\frac{\hat{W}_q^{(K)}}{K} = \frac{1}{2} + \frac{1}{2K(1-\rho)} - \frac{1}{2K} \frac{\sum_{j=0}^{K} \frac{(K\rho)^j}{j!}(K-j+1)(K+j)}{\sum_{j=0}^{K-1} \frac{(K\rho)^j}{j!}(K-j)} - \rho \quad (60)$$

(par (30) et (56)).

$$\text{et } \lim_{K\to\infty} \frac{\hat{W}_q^{(K)}}{K} = \frac{1}{2} - \frac{(1-\rho)}{2} = \frac{\rho}{2} \quad (61)$$

(par (30) et (57)).

Note : $\lim_{K\to\infty} \frac{\hat{W}_q^{(K)}}{K} = \lim_{K\to\infty} \frac{W_q^{(K)}}{K}$ pour les quasi-files Q-M/$D^{[K]}$/1.

Pour avoir une idée de la "qualité" de cette approximation, nous avons calculé $\frac{\tilde{W}^{(K)}}{K}$ et nous l'avons comparé à la "valeur exacte" de de $\frac{W_q^{(K)}}{K}$ pour un certain nombre de valeurs des paramètres. Dans le tableau 5.3, les "valeurs exactes" de $\frac{W_q^{(K)}}{K}$ sont tirées de Downton (1955, 1956) ; elles ont été obtenues en utilisant les résultats des théorèmes 5.3 et 5.4. Les "valeurs exactes" du tableau 5.4 ont été déduites des tables de Hillier et Yu (1981) pour les files d'attente $E_K/E_k/1$. Nous remercions Jacques Roy qui nous a fourni les "valeurs exactes" du tableau 5.5.

Un examen rapide des tableaux 5.3, 5.4 et 5.5 nous permet de constater que l'approximation proposée fournit d'excellents résultats, et ce, même pour des valeurs très différentes des paramètres K et k. Il semble donc tout à fait raisonnable de s'en contenter quand on n'a que quelques temps d'attente à évaluer. Pour ce qui est de l'utilisation directe de cette approximation dans le cadre d'un modèle complexe, nous verrons dans les sous-sections qui suivent qu'on peut définir une approche plus efficace.

Il faut remarquer dans les tableaux que, quand K est égal à 1, on a un accord parfait entre les valeurs approchées et les valeurs exactes. Cela n'est pas l'effet du hasard. En effet, quand K=1, la file d'attente associée à la quasi-file qui nous intéresse est de type $M/E_k/1$ ou $M/D/1$. Dans les deux cas, les intervalles interarrivées suivent donc une loi exponentielle. Le caractère "sans mémoire" de l'exponentielle fait que la durée des périodes creuses (c'est-à-dire I) suit la même loi que la durée des intervalles interarrivées, et que, si l'on distingue les

K	valeur	ρ = 0,9				ρ = 0,8			
		k=1	k=2	k=3	k=∞	k=1	k=2	k=3	k=∞
1	approchée	9,00	6,75	6,00	4,50	4,00	3,00	2,670	2,000
	exacte	9,00	6,75	6,00	4,50	4,00	3,00	2,670	2,000
2	approchée	6,81	4,58	3,84	2,37	3,06	2,08	1,750	1,111
	exacte	6,83	4,60	3,86	2,38	3,08	2,09	1,770	1,123
3	approchée	6,08	3,86	3,13	1,68	2,75	1,78	1,460	0,831
	exacte	6,11	3,88	3,15	1,69	2,77	1,79	1,470	0,843
6	approchée	5,36	3,15	2,42	1,005	2,44	1,48	1,167	0,573
	exacte	5,39	3,17	2,44	1,017	2,46	1,50	1,182	0,581
12	approchée	5,00	2,80	2,08	0,690	2,29	1,33	1,026	0,461
	exacte	5,02	2,82	2,09	0,698	2,31	1,35	1,041	0,467
24	approchée	4,82	2,62	1,90	0,547	2,21	1,26	0,957	0,418
	exacte	4,84	2,64	1,92	0,552	2,23	1,28	0,971	0,421
∞	approchée	4,64	2,45	1,73	0,450	2,14	1,19	0,889	0,400
	exacte	4,66	2,47	1,75	0,450	2,15	1,21	0,902	0,400

Tableau 5.3 : Comparaison des résultats obtenus par l'approximation de base avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/1 et Q-M/$D^{[K]}$/1 avec taux d'arrivée unitaire.

K	valeur	ρ = 0,5			ρ = 0,99		
		k=1	k=2	k=3	k=1	k=2	k=3
1	approchée	1,000	0,750	0,667	99,00	74,25	66,00
	exacte	1,000	0,750	0,667	99,00	74,25	66,00
2	approchée	0,800	0,562	0,485	74,31	49,58	41,34
	exacte	0,809	0,570	0,492	74,33	49,60	41,36
3	approchée	0,737	0,506	0,431	66,09	41,37	33,13
	exacte	0,747	0,514	0,438	66,11	41,39	33,16
4	approchée	0,706	0,479	0,407	61,97	37,26	29,03
	exacte	0,716	0,487	0,413	62,00	37,28	29,05
9	approchée	0,657	0,438	0,371	55,12	30,42	22,20
	exacte	0,666	0,445	0,375	55,15	30,44	22,22
16	approchée	0,640	0,425	0,360	52,72	28,03	19,81
	exacte	0,649	0,431	0,364	52,77	28,05	19,83

Tableau 5.4 : Comparaison des résultats obtenus par l'approximation de base avec ceux déduits des tables de Hillier et Yu pour le temps d'attente dans les quasi-files d'attente $Q\text{-}M/E_k^{[K]}/1$ avec taux d'arrivée unitaire.

K	valeur	ρ=0,80	ρ=0,90	ρ=0,92	ρ=0,94	ρ=0,96	ρ=0,98
6	approchée	0,573	1,005	1,216	1,566	2,264	4,350
	exacte	0,581	1,017	1,228	1,580	2,277	4,364
12	approchée	0,461	0,690	0,798	0,975	1,325	2,370
	exacte	0,467	0,698	0,807	0,985	1,336	2,381
24	approchée	0,418	0,547	0,603	0,694	0,871	1,396
	exacte	0,421	0,552	0,609	0,700	0,879	1,404
50	approchée	0,403	0,483	0,513	0,559	0,647	0,901
	exacte	0,404	0,486	0,516	0,563	0,651	0,906
100	approchée	0,400	0,460	0,478	0,504	0,550	0,680
	exacte	0,401	0,461	0,480	0,506	0,553	0,684
150	approchée	0,400	0,454	0,469	0,488	0,521	0,609
	exacte	0,400	0,455	0,470	0,490	0,523	0,612
200	approchée	0,400	0,452	0,465	0,481	0,507	0,575
	exacte	0,400	0,453	0,466	0,482	0,509	0,577

Tableau 5.5 : Comparaison des résultats obtenus par l'approximation de base avec ceux de Roy pour le temps d'attente dans les quasi-files d'attente $Q\text{-}M/D^{[K]}/1$ avec taux d'arrivée unitaire.

périodes creuses en fonction du nombre d'intervalles de service complétés dans l'intervalle interarrivées qui les engendre, on a dans tous les cas la même distribution. C'est donc dire que, bien que correspondant à des processus distincts, les V.A. I et Z ont la même distribution. Qui plus est, on peut affirmer que cette loi commune est une exponentielle de moyenne 1 (distribution des intervalles interarrivées), ce qui nous permet de conclure :

$$\frac{E[Z^2]}{2E[Z]} = \frac{E[I^2]}{2E[I]} = 1 \ .$$

5.7.5 Approximation hyperbolique sur la taille des groupes

Le calcul du terme $\frac{E[Z^2]}{2E[Z]}$ de l'expression donnant $\frac{\hat{W}_q^{(K)}}{K}$ peut se révéler assez fastidieuse surtout quand K est grand. D'autre part, si l'on veut utiliser l'approximation de base pour évaluer (avec une certaine erreur, il est vrai) le temps d'attente dans des quasi-files avec service en groupes de tailles variables en prenant comme valeur de K la moyenne de la taille des groupes servis, des problèmes se poseront si cette moyenne n'est pas un nombre entier. Il paraît donc important de définir une formule d'approximation modifiée qui permette de contourner ces difficultés.

Pour k, ρ et K donnés, appelons $\tilde{Z}(k,\rho,K)$ la valeur de $\frac{E[Z^2]}{2E[Z]}$ pour la file d'attente $E_K/E_k/1$ avec coefficient d'achalandage ρ et taux d'arrivée unitaire (par convention $E_\infty \equiv D$ dans la suite). Si pour k et ρ donnés, on observe la suite formée par les valeurs de $\tilde{Z}(k,\rho,K)$ quand K varie de 1 à l'infini, on remarque que c'est une suite monotone strictement décroissante dont le premier terme ($\tilde{Z}(k,\rho,1)$) est toujours égal à 1

(voir remarque à la fin de la sous-section précédente), et dont la limite à l'infini ($\tilde{Z}(k,\rho,\infty)$) est toujours strictement positive (voir tableau 5.6). D'autre part, on sait que pour K=1, $\tilde{Z}(k,\rho,1)$ correspond à la véritable valeur de $\frac{E[I^2]}{2E[I]}$. Ceci nous amène à suggérer de remplacer, dans nos formules d'approximation de $\frac{W_q^{(K)}}{K}$, $\tilde{Z}(k,\rho,K)$ par l'approximation "hyperbolique" suivante :

$$Z^h(k,\rho,K) = \frac{1}{K}\tilde{Z}(k,\rho,1) + \frac{K-1}{K}Z(k,\rho,\infty)$$

$$\Rightarrow \quad Z^h(k,\rho,K) = \frac{1}{K} + \frac{K-1}{K}\tilde{Z}(k,\rho,\infty) \qquad (62)$$

Les deux propriétés les plus importantes de cette approximation hyperbolique sont :

- $Z^h(k,\rho,1) = \tilde{Z}(k,\rho,1) = 1$
- $\lim_{K\to\infty} Z^h(k,\rho,K) = \tilde{Z}(k,\rho,\infty)$

Le tableau 5.6 donne les valeurs de $Z^h(k,\rho,K)$ et de $\tilde{Z}(k,\rho,K)$ pour un certain nombre de valeurs choisies des paramètres.

Si l'on remplace $\tilde{Z}(k,\rho,K)$ par $Z^h(k,\rho,K)$ dans les formules pour $\frac{\hat{W}_q^{(K)}}{K}$, on obtiendra une nouvelle approximation de $\frac{W_q^{(K)}}{K}$ que l'on appellera l'approximation "hyperbolique" de $\frac{W_q^{(K)}}{K}$ et que l'on dénotera $\frac{\hat{W}_q^{(K)h}}{K}$.

On a pour les quasi-files Q-M/$E_k^{[K]}$/1

K	valeur	ρ = 0,9				ρ = 0,8			
		k=1	k=2	k=3	k=∞	k=1	k=2	k=3	k=∞
1	$Z(k,\rho,K)$	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	$Z^h(k,\rho,K)$	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
2	$Z(k,\rho,K)$	0,689	0,670	0,660	0,631	0,691	0,673	0,664	0,639
	$Z^h(k,\rho,K)$	0,681	0,652	0,635	0,525	0,683	0,655	0,639	0,550
3	$Z(k,\rho,K)$	0,582	0,553	0,538	0,491	0,585	0,558	0,544	0,502
	$Z^h(k,\rho,K)$	0,575	0,536	0,513	0,367	0,577	0,540	0,519	0,400
6	$Z(k,\rho,K)$	0,473	0,432	0,410	0,328	0,476	0,438	0,417	0,344
	$Z^h(k,\rho,K)$	0,468	0,420	0,391	0,208	0,471	0,425	0,398	0,250
12	$Z(k,\rho,K)$	0,418	0,369	0,341	0,227	0,421	0,375	0,349	0,247
	$Z^h(k,\rho,K)$	0,415	0,362	0,330	0,129	0,418	0,368	0,338	0,175
24	$Z(k,\rho,K)$	0,390	0,337	0,306	0,162	0,393	0,343	0,314	0,186
	$Z^h(k,\rho,K)$	0,389	0,333	0,299	0,090	0,391	0,339	0,308	0,138
100	$Z(k,\rho,K)$	0,369	0,312	0,278	0,090	0,372	0,318	0,287	0,125
	$Z^h(k,\rho,K)$	0,368	0,311	0,276	0,060	0,371	0,317	0,285	0,109
∞	$Z(k,\rho,K)$	0,362	0,304	0,269	0,050	0,365	0,310	0,278	0,100
	$Z^h(k,\rho,K)$	0,362	0,304	0,269	0,050	0,365	0,310	0,278	0,100

Tableau 5.6 : Valeurs de $Z(k,\rho,K)$ et $Z^h(k,\rho,K)$ pour $k=1,2,3,\infty$, $K=1,2,3,6,12,24,100,\infty$ et $\rho=0,8$ et $0,9$.

$$\frac{\hat{W}_q^{(K)h}}{K} = \frac{1}{2} + \frac{(1/K)+(\rho/k)}{2(1-\rho)} - [\frac{1}{K} + \frac{K-1}{K} Z(k,\rho,\infty)] \tag{63}$$

$$\text{et} \quad \lim_{K\to\infty} \frac{\hat{W}_q^{(K)h}}{K} = \frac{1}{2} + \frac{\rho}{2k(1-\rho)} - \tilde{Z}(k,\rho,\infty) \tag{64}$$

$$\text{avec} \quad \tilde{Z}(k,\rho,\infty) = 1 - \frac{1}{2} \frac{1-(\frac{\rho}{k})^2 k(k+1)+(\frac{\rho}{k})^2 e^{-k/\rho} \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!}(k-j+1)(k+j)}{1-\rho+(\frac{\rho}{k})e^{-k/\rho} \sum_{j=0}^{k} \frac{(k/\rho)^j}{j!}(k-j)}$$

Pour les quasi-files Q-M/D$^{[K]}$/1, on obtient

$$\frac{\hat{W}_q^{(K)h}}{K} = \frac{1}{2} + \frac{1}{2K(1-\rho)} - [\frac{1}{K} + \frac{K-1}{K} \frac{1-\rho}{2}]$$

$$\Rightarrow \quad \frac{\hat{W}_q^{(K)h}}{K} = \frac{\rho}{2} + \frac{\rho^2}{2K(1-\rho)} \tag{65}$$

$$\text{et} \quad \lim_{K\to\infty} \frac{W_q^{(K)h}}{K} = \frac{\rho}{2} \tag{66}$$

Le tableau 5.7 permet de comparer les résultats obtenus avec l'approximation hyperbolique à ceux de Downton déjà utilisés dans le tableau 5.3 (note : pour K=1 et K=∞, les résultats obtenus sont identiques à ceux indiqués dans le tableau 5.3).

Si l'on examine les résultats du tableau 5.7, on ne peut qu'être frappé par la marge extrêmement faible existant entre les valeurs approchées et les valeurs exactes quand k est petit. En fait, dans ce cas, l'approximation hyperbolique est supérieure à l'approximation de base, ce qui est quelque peu surprenant. A l'infini, cependant, les résultats

K	valeur	ρ = 0,9				ρ = 0,8			
		k=1	k=2	k=3	k=∞	k=1	k=2	k=3	k=∞
2	approchée	6,82	4,60	3,87	2,48	3,07	2,10	1,78	1,200
	exacte	6,83	4,60	3,86	2,38	3,08	2,09	1,77	1,123
3	approchée	6,09	3,88	3,15	1,80	2,76	1,79	1,48	0,933
	exacte	6,11	3,88	3,15	1,69	2,77	1,79	1,47	0,843
6	approchée	5,37	3,16	2,44	1,125	2,45	1,49	1,186	0,667
	exacte	5,39	3,17	2,44	1,017	2,46	1,50	1,182	0,581
12	approchée	5,00	2,81	2,09	0,787	2,29	1,34	0,037	0,533
	exacte	5,02	2,82	2,09	0,698	2,31	1,35	1,041	0,467
24	approchée	4,82	2,63	1,91	0,619	2,21	1,27	0,963	0,467
	exacte	4,84	2,64	1,92	0,552	2,23	1,28	0,971	0,421

Tableau 5.7 : Comparaison des résultats obtenus par l'approximation hyperbolique avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/ et Q-M/$D^{[K]}$/1 avec taux d'arrivée unitaire.

obtenus en utilisant l'approximation hyperbolique sont nettement plus décevants puisque l'erreur relative d'approximation peut atteindre 15% (pour ρ=0,8 et K=12) ; ceci semblerait donc indiquer que la performance de cette nouvelle approximation se dégrade quand le paramètre k augmente.

5.7.6 Formule d'approximation simplifiée

Avec la formule d'approximation hyperbolique, nous sommes forcés de calculer $\tilde{Z}(k,\rho,\infty)$ chaque fois que l'on veut estimer $\frac{W_q(K)}{K}$. Cela nous causera des difficultés quand nous voudrons utiliser cette approximation dans le modèle d'équilibre, car, si k est un paramètre fixe pour chaque arrêt d'autobus, ρ ne l'est pas, c'est une quantité définie par rapport aux variables du modèle. L'idéal pour nous consisterait à trouver, pour chaque valeur de k, une fonction relativement "simple" de ρ qui permette d'approximer adéquatement $\tilde{Z}(k,\rho,\infty)$. En reformulant légèrement cette problématique, nous dirons que nous cherchons une fonction $Z_\infty^S(k,\rho)$ telle que

$$Z_\infty^S(k,\rho) \simeq \tilde{Z}(k,\rho,\infty) \ .$$

La première étape dans notre recherche consiste en un calcul de $\tilde{Z}(k,\rho,\infty)$ pour un nombre suffisant de valeur de k et de ρ(voir tableau 5.8). Si l'on observe $\tilde{Z}(k,\rho,\infty)$, pour k donné, en tant que fonction de ρ, on peut remarquer que $\tilde{Z}(k,0,\infty)$ = 0,05 (quel que soit k), et que $\tilde{Z}(k,\rho,\infty)$ semble se comporter comme une fonction monotone strictement décroissante quand ρ augmente de 0 à 1. On sait aussi que, quand k tend vers l'infini, $\tilde{Z}(k,\rho,\infty)$ tend vers $(1-\rho)/2$; il pourrait donc être intéressant d'étudier le comportement du terme "résiduel" $\tilde{Z}(k,\rho,\infty)-(1-\rho)/2$. Ce terme varie, pour k donné, de 0 à $\tilde{Z}(k,1,\infty)$ quand ρ varie de 0 à 1. On peut donc penser à

une interpolation quelconque de ce terme entre 0 et 1. Cette interpolation connue, il suffira de lui ajouter $(1-\rho)/2$ pour obtenir $Z_\infty^S(k,\rho)$. Par un processus de tâtonnements, nous sommes arrivés assez laborieusement à la formule suivante :

$$Z_\infty^S(k,\rho) = \frac{1-\rho}{2} + \frac{\rho^2}{2}\left[\left(Z_k^* + \frac{1}{k}\right) + \left(Z_k^* - \frac{1}{k}\right)\rho^2\right]$$

où $\quad Z_k^* \equiv \tilde{Z}(k,1,\infty)$.

Le tableau 5.8 permet de comparer $Z_\infty^S(k,\rho)$ et $\tilde{Z}(k,\rho,\infty)$ pour plusieurs valeurs des paramètres. Comme on peut le voir, pour k assez petit, $Z_\infty^S(k,\rho)$ est très proche de $\tilde{Z}(k,\rho,\infty)$.

Si, dans l'approximation hyperbolique de $\tilde{Z}(k,\rho,K)$, on remplace $\tilde{Z}(k,\rho,\infty)$ par $Z_\infty^S(k,\rho)$, on obtient une nouvelle approximation de $\tilde{Z}(k,\rho,K)$ que nous appellerons l'"approximation simplifiée" :

$$Z^S(k,\rho,K) = \frac{1}{K} + \frac{K-1}{K}\left\{\frac{1-\rho}{2} + \frac{\rho^2}{2}\left[\left(Z_k^* + \frac{1}{k}\right) + \left(Z_k^* - \frac{1}{k}\right)\rho^2\right]\right\} .$$

De cette approximation de $\tilde{Z}(k,\rho,K)$, nous tirons l'"approximation simplifiée" de $\frac{W_q^{(K)}}{K}$ pour les quasi-files d'attente Q-M/$E_k^{[K]}$/1 :

$$\frac{\hat{W}_q^{(K)S}}{K} = \frac{1}{2} + \frac{(1/K)+(\rho/k)}{2(1-\rho)} - \left\{\frac{1}{K} + \frac{K-1}{K}\left\{\frac{1-\rho}{2} + \frac{\rho^2}{2}\left[\left(Z_k^* + \frac{1}{k}\right) + \left(Z_k^* - \frac{1}{k}\right)\rho^2\right]\right\}\right\} \quad (67)$$

où $$Z_k^* = 1 - \frac{1}{2}\,\frac{1 - \frac{(k+1)}{k} + \frac{1}{k^2}\,e^{-k}\sum_{j=0}^{k}\frac{k^j}{j!}(k-j+1)(k+j)}{\frac{1}{k}\,e^{-k}\sum_{j=0}^{k-1}\frac{k^j}{j!}(k-j)} .$$

ρ	k=1		k=2		k=3		k=4		k=50	
	$Z(k,\rho,\infty)$	$Z_\infty^S(k,\rho)$	$Z(k,\rho,\infty)$	$Z_\infty^S(k,\rho)$	$Z(k,\rho,\infty)$	$Z_\infty^S(k,\rho)$	$Z(k,\rho,\infty)$	$Z_\infty^S(k,\rho)$	$Z(k,\rho,\infty)$	$Z_\infty^S(k,\rho)$
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
0,1	0,456	0,457	0,453	0,454	0,452	0,453	0,451	0,452	0,450	0,451
0,2	0,424	0,425	0,412	0,415	0,408	0,412	0,406	0,410	0,401	0,402
0,3	0,404	0,403	0,381	0,383	0,371	0,376	0,366	0,372	0,351	0,355
0,4	0,390	0,388	0,357	0,358	0,342	0,345	0,332	0,339	0,303	0,310
0,5	0,381	0,380	0,340	0,337	0,319	0,320	0,305	0,310	0,255	0,267
0,6	0,374	0,375	0,327	0,322	0,302	0,300	0,285	0,286	0,206	0,225
0,7	0,369	0,373	0,318	0,311	0,288	0,284	0,268	0,267	0,166	0,186
0,8	0,365	0,371	0,310	0,304	0,278	0,272	0,256	0,253	0,103	0,149
0,8	0,362	0,367	0,304	0,300	0,269	0,265	0,246	0,243	0,102	0,114
1,0	0,359	0,359	0,299	0,299	0,262	0,262	0,237	0,237	0,082	0,082

Tableau 5.8 : Valeurs de $Z(k,\rho,\infty)$ et $Z_\infty^S(k,\rho)$ pour k=1, 2, 3, 4, 50 et ρ variant de 0,0 à 1,0 par pas de 0,1.

Après simplifications, on peut réécrire Z_k^* sous la forme

$$Z_k^* = 1 - \frac{1}{2} \frac{-e^k + \frac{1}{k} \sum_{j=0}^{k} \frac{k^j}{j!} (k-j+1)(k+j)}{\sum_{j=0}^{k-1} \frac{k^j}{j!} (k-j)} \quad . \tag{68}$$

De (67), on peut tirer une expression pour $\hat{W}_q^{(K)S}$:

$$\hat{W}_q^{(K)S} = \frac{K}{2} + \frac{1}{2(1-\rho)} + \frac{K\rho}{2k(1-\rho)} - 1 - (K-1)\{\frac{1-\rho}{2} + \frac{\rho^2}{2}[(Z_k^* + \frac{1}{k}) + (Z_k^* - \frac{1}{k})\rho^2]\} \tag{69}$$

$$= \frac{K\rho}{2} + \frac{K\rho}{2k(1-\rho)} + \frac{1}{2(1-\rho)} - \frac{1+\rho}{2} - \frac{\rho^2}{2} (K-1)[(Z_k^* + \frac{1}{k}) + (Z_k^* - \frac{1}{k})\rho^2]$$

$$\hat{W}_q^{(K)S} = \frac{K\rho}{2}[1 + \frac{1}{k(1-\rho)}] + \frac{\rho^2}{2(1-\rho)} - \frac{\rho^2}{2}(K-1)[(Z_k^* + \frac{1}{k}) + (Z_k^* - \frac{1}{k})\rho^2] \tag{70}$$

Le tableau 5.9 permet de comparer les valeurs obtenues en ulitisant l'approximation simplifiée aux valeurs exactes de Downton. Comme l'on pouvait s'y attendre après avoir vu le tableau 5.8, l'approximation est des plus satisfaisantes.

Avant de clore cette section, il serait bon de se rappeler que, dans tous nos calculs, nous avons supposé que le taux d'arrivée λ était égal à 1. Si λ est différent de 1, il faut diviser par λ la valeur donnée par (70) pour obtenir $\hat{W}_q^{(K)S}$. On a donc

$$\hat{W}_q^{(K)S} = \frac{1}{\lambda}\{\frac{K\rho}{2}[1 + \frac{1}{k(1-\rho)}] + \frac{\rho^2}{2(1-\rho)} - \frac{\rho^2}{2}(K-1)[(Z_k^* + \frac{1}{k}) + (Z_k^* - \frac{1}{k})\rho^2]\}$$

$$= \frac{K\rho}{2\lambda}\{[1 + \frac{1}{k(1-\rho)}] + \frac{\rho}{K(1-\rho)} - \rho(1 - \frac{1}{K})[(Z_k^* + \frac{1}{k}) + (Z_k^* - \frac{1}{k})\rho^2]\}$$

K	valeur	ρ = 0,9			ρ = 0,8		
		k=1	k=2	k=3	k=1	k=2	k=3
1	approchée	9,00	6,75	6,00	4,00	3,00	2,67
	exacte	9,00	6,75	6,00	4,00	3,00	2,67
2	approchée	6,82	4,60	3,87	3,06	2,10	1,78
	exacte	6,83	4,60	3,86	3,08	2,09	1,77
3	approchée	6,08	3,88	3,16	2,75	1,80	1,49
	exacte	6,11	3,88	3,15	2,77	1,79	1,47
6	approchée	5,36	3,17	2,45	2,44	1,50	1,191
	exacte	5,39	3,17	2,44	2,46	1,50	1,182
12	approchée	5,00	2,81	2,09	2,28	1,35	1,042
	exacte	5,02	2,82	2,09	2,31	1,35	1,041
24	approchée	4,81	2,63	1,91	2,21	1,27	0,969
	exacte	4,84	2,64	1,92	2,23	1,28	0,971
∞	approchée	4,63	2,45	1,74	2,13	1,20	0,895
	exacte	4,66	2,47	1,75	2,15	1,21	0,902

Tableau 5.9 : Comparaison des résultats obtenus par l'approximation simplifiée avec ceux de Downton pour le temps d'attente dans les quasi-files d'attente Q-M/$E_k^{[K]}$/1 avec taux d'arrivée unitaire.

$$\Rightarrow \quad \hat{W}_q^{(K)S} = \frac{1}{2\mu}\left\{\left[1 + \frac{1}{k(1-\rho)}\right] + \frac{\rho}{K(1-\rho)} - \rho\left(1 - \frac{1}{K}\right)\left[\left(Z_k^* + \frac{1}{k}\right) + \left(Z_k^* - \frac{1}{k}\right)\rho^2\right]\right\}. \qquad (71)$$

en utilisant l'identité $\rho = \frac{\lambda}{K\mu}$.

5.8 Fonctions d'attente

Pour définir les fonctions d'attente du modèle d'équilibre, nous allons tout simplement utiliser la relation (71) en faisant les identifications nécessaires entre les paramètres de (71) et les variables et paramètres du modèle qui leur correspondent.

Pour un arrêt r donné, appelons $\mu_{\ell(r)}$ la fréquence moyenne d'autobus sur la ligne desservant l'arrêt r et k(r) le paramètre de forme de la distribution des intervalles interautobus à l'arrêt r ; on peut établir immédiatement les correspondances $\mu \leftrightarrow \mu_{\ell(r)}$ et $k \leftrightarrow k(r)$.

Essayons maintenant de déterminer la "capacité effective" disponible à l'arrêt r. La capacité totale de la ligne desservant l'arrêt est $K_{\ell(r)}$, mais il y a en moyenne $\bar{v}_r$ places déjà occupées dans les autobus ; la "capacité effective" de la ligne est donc de $(K_{\ell(r)} - \bar{v}_r)$ places par unité de temps et la capacité effective par autobus est égale à $(K_{\ell(r)} - \bar{v}_r)/\mu_{\ell(r)}$. Sachant, d'autre part, que le taux moyen d'arrivée des usagers à l'arrêt est $\tilde{v}_r$, on peut en déduire que le coefficient d'achalandage ρ est égal à $\dfrac{\tilde{v}_r}{K_{\ell(r)} - \bar{v}_r}$. Le tableau 5.10 résume les correspondances que nous venons d'établir.

Paramètres du modèle d'attente	Variables et paramètres du modèle d'équilibre
μ	$\mu_{\ell(r)}$
k	$k(r)$
K	$(K_{\ell(r)} - \bar{v}_r)/\mu_{\ell(r)}$
ρ	$\dfrac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r}$

Tableau 5.10 : Correspondance entre les paramètres du modèle d'attente et les variables et paramètres du modèle d'équilibre pour un arrêt simple r donné.

Pour déterminer $\Phi_r(\bar{v}_r,\tilde{v}_r)$ la fonction d'attente à l'arrêt r, il ne nous reste plus qu'à remplacer dans (71) μ, k, K et ρ par leur équivalent dans le modèle d'équilibre. On obtient

$$\Phi_r(\bar{v}_r,\tilde{v}_r) = \frac{1}{2\mu_{\ell(r)}} \{ [1 + \frac{1}{k(r)(1 - \frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r})}]$$

$$+ \frac{\frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r}}{\frac{(K_{\ell(r)}-\bar{v}_r)}{\mu_{\ell(r)}}(1 - \frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r})}$$

$$- (\frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r})(1 - \frac{1}{\frac{K_{\ell(r)}-\bar{v}_r}{\mu_{\ell(r)}}}) [Z^*_{k(r)} + \frac{1}{k(r)}) + (Z^*_{k(r)} - \frac{1}{k(r)})$$

$$(\frac{\tilde{v}_r}{K_{\ell(r)}-\mu_{\ell(r)}})^2] \}$$

Après quelques simplifications, on arrive à

$$\Phi_r(\bar{v}_r,\tilde{v}_r) = \frac{1}{2\mu_{\ell(r)}} \{[\,1 + \frac{K_{\ell(r)}-\bar{v}_r}{k(r)(K_{\ell(r)}-\bar{v}_r-\tilde{v}_r)}\,] + \frac{\mu_{\ell(r)}\,\tilde{v}_r}{(K_{\ell(r)}-\bar{v}_r)(K_{\ell(r)}-\bar{v}_r-\tilde{v}_r)}$$

$$- \frac{\tilde{v}_r(K_{\ell(r)}-\bar{v}_r-\mu_{\ell(r)})}{(K_{\ell(r)}-\bar{v}_r)^2}\,[\,(Z^*_{k(r)} + \frac{1}{k(r)}) + (Z^*_{k(r)} - \frac{1}{k(r)})$$

$$(\frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r})^2\,]\,\} \qquad (72)$$

Pour le modèle d'équilibre, $(Z^*_{k(r)} + \frac{1}{k(r)})$ et $(Z^*_{k(r)} - \frac{1}{k(r)})$ sont des constantes qui peuvent être calculées dans une phase d'initialisation. On peut donc poser

$$\zeta_k^+ = Z^*_{k(r)} + \frac{1}{k(r)}$$

et $$\zeta^- = Z^*_{k(r)} - \frac{1}{k(r)}\ .$$

(72) devient alors

$$\Phi_r(\bar{v}_r,\tilde{v}_r) = \frac{1}{2\mu_{\ell(r)}} \{[\,1 + \frac{K_{\ell(r)}-\bar{v}_r}{k(r)(K_{\ell(r)}-\bar{v}_r-\tilde{v}_r)}\,] + \frac{\mu_{\ell(r)}\tilde{v}_r}{(K_{\ell(r)}-\bar{v}_r)(K_{\ell(r)}-\bar{v}_r-\tilde{v}_r)}$$

$$- \frac{\tilde{v}_r(K_{\ell(r)}-\bar{v}_r-\mu_{\ell(r)})}{(K_{\ell(r)}-\bar{v}_r)^2}\,[\,\zeta_k^+ + \zeta_k^-(\frac{\tilde{v}_r}{K_{\ell(r)}-\bar{v}_r})^2\,]\,\} \qquad (73)$$

où $$\zeta_k^+ = \frac{1}{k(r)} + 1 - \frac{1}{2}\,\frac{-e^{k(r)} + \frac{1}{k(r)} \sum_{j=0}^{k(r)} \frac{[k(r)]^j}{j!}\,[k(r)-j+1][k(r)+j]}{k(r)-1 \sum_{j=0} \frac{[k(r)]^j}{j!}\,[k(r)-j]}$$

et $\zeta_k^- = \zeta_k^+ - \frac{2}{k(r)}$.

5.9 Critiques des hypothèses sous-jacentes au modèle

Dans notre élaboration de $\Phi_r(\bar{v}_r, \tilde{v}_r)$, nous avons fait, plus ou moins implicitement, trois hypothèses importantes qui méritent d'être discutées.

La plus importante de ces hypothèses consiste à utiliser pour définir les fonctions d'attente le temps d'attente dans des quasi-files en régime stationnaire. Il paraît certain que le phénomène d'attente à un arrêt donné pendant une période de pointe constitue un phénomène transitoire dans lequel certains paramètres (taux d'arrivée des usagers, nombre moyen de places occupées dans les autobus) varient dans le temps.

Trois raisons militent cependant en faveur de l'utilisation des temps d'attente en régime stationnaire. La première et la plus évid de ces trois raisons est liée à l'état actuel des connaissances dans le domaine des systèmes d'attente en régime transitoire : la détermination du temps d'attente moyen dans un système d'attente dont les paramètres varient dynamiquement est un problème d'une très grande complexité et il paraît peu probable que l'on puisse le résoudre de manière satisfaisante dans un avenir rapproché. La seconde raison est aussi de nature pratique : à supposer que l'on réussisse à mettre au point une méthode permettant de calculer le temps d'attente dans une quasi-file d'attente en régime transitoire, sera-t-il possible de rassembler les données nécessaires à l'emploi d'une telle méthode (fonctions donnant le taux d'arrivée et les autres

paramètres du problème dans le temps)? Chose certaine, il sera beaucoup plus difficile de "calibrer" le modèle dans un tel cas. La troisième raison, quant à elle, est plutôt "philosophique" : il paraît plus raisonnable d'utiliser, dans le cadre d'un modèle d'équilibre statique, un sous-modèle d'arrêt en régime stationnaire (dans lequel les temps d'attente moyens sont invariants dans le temps) qu'un sous-modèle d'arrêt en régime transitoire. Cette question de compatibilité entre la nature du modèle global et celle du sous-modèle d'arrêt semble assez fondamentale ; dans le cas où les résultats produits par le modèle global d'équilibre avec le sous-modèle d'arrêt en régime stationnaire ne seraient pas satisfaisants, c'est probablement toute la méthodologie qui devrait être remise en question, et il faudrait passer à l'élaboration de modèles dynamiques pour l'affectation des passagers. En résumé, il faut donc en venir à la conclusion que, bien que l'hypothèse de stationnarité ne puisse être justifiée a priori, elle est tout à fait nécessaire au calcul des fonctions d'attente dans l'état actuel des connaissances.

La seconde hypothèse se situe au niveau du processus des arrivées des usagers. Nous avons supposé que les arrivées suivaient un processus de Poisson et qu'elles étaient indépendantes du mécanisme de service. Il nous semble inutile d'insister sur la question de l'indépendance entre les arrivées des usagers et les passages d'autobus à l'arrêt puisque nous en avons déjà discuté dans la sous-sections 3.8.4. Pour ce qui est du choix de la loi de Poisson, il nous paraît justifiable pour les arrêts auxquels on ne peut faire de correspondances. Pour les points de correspondances, il faudrait peut-être considérer un processus avec arrivées en

groupes, mais il n'est pas certain que les complications additionnelles que cela impliquerait dans le modèle en vaudraient la peine.

La dernière hypothèse consiste à utiliser pour une quasi-file avec service en groupes de taille variable les résultats relatifs aux quasi-files avec service en groupes de taille fixe. Si l'on se fie à certains résultats très fragmentaires de simulations que nous avons effectuées (voir chapitre huitième), l'erreur commise en procédant ainsi est très minime.

En conclusion, il faut se rappeler que, même si l'on ne peut affirmer avec certitude que les fonctions d'attente que nous avons définies dans ce chapitre sont adéquates, elles nous permettent d'utiliser le modèle d'équilibre.

CHAPITRE SIXIÈME

FORMULATION DU MODÈLE D'AFFECTATION D'ÉQUILIBRE POUR LES RÉSEAUX COMPORTANT DES ARRÊTS MULTIPLES

6.1 Introduction

Dans le chapitre quatrième, nous avons formulé un modèle d'équilibre pour l'affectation des passagers en nous restreignant à ne considérer que les réseaux sans arrêts multiples; c'est une restriction qui limite énormément les possibilités d'utilisation du modèle, car, en pratique, la plupart des réseaux de transport en commun comportent des arrêts multiples. Evidemment, il serait toujours possible d'utiliser le modèle du chapitre quatrième pour de tels réseaux en remplaçant les arrêts multiples par plusieurs arrêts simples, mais on déformerait le problème, puisque, comme nous l'avons indiqué dans l'introduction, les usagers qui se présentent aux arrêts multiples peuvent attendre les autobus de plusieurs lignes à la fois. Il est donc nécessaire de modifier la formulation du modèle d'équilibre pour tenir explicitement compte des arrêts multiples. Pour ce faire, nous devrons, dans un premier temps, préciser le comportement des usagers aux arrêts multiples et caractériser les stratégies découlant de ces hypothèses supplémentaires de comportement. Cela nous amènera par la suite à définir un sous-modèle d'arrêt multiple (qui sera étudié en détail au chapitre septième), puis à développer un nouvel ensemble de conditions pour les affectations réalisables et les affectations d'équilibre par un cheminement analogue à celui du chapitre quatrième.

6.2 Définitions préliminaires, notation additionnelle

Avant de préciser le comportement des usagers aux arrêts multiples, il est essentiel de nous munir d'un certain nombre de définitions et de notations pour le réseau augmenté. On peut reprendre pour celui-ci presque en totalité les définitions et la notation du chapitre quatrième; il faut cependant souligner un certain nombre d'ajouts et de modifications importants.

Pour un arrêt multiple, on ne parle pas de son prédécesseur ou de son successeur, car ceux-ci ne sont pas en général uniques, mais plutôt de son précesseur ou de son successeur par une ligne donnée (on peut bien sûr parler aussi de l'ensemble des successeurs ou des précédesseurs d'un arrêt multiple).

La notation additionnelle que nous utiliserons par la suite est la suivante :

R^s, ensemble des arrêts simples;

R^m, ensemble des arrêts multiples;

L_r, ensemble des lignes passant à l'arrêt multiple r;

L_r^+, ensemble des lignes passant à l'arrêt multiple r et pour lesquelles r n'est pas un terminus d'arrivée;

L_r^-, ensemble des lignes passant à l'arrêt multiple r et pour lesquelles r n'est pas un terminus de départ;

$\gamma_\ell^-(r)$, précédesseur par la ligne ℓ de l'arrêt multiple r;

$\gamma_\ell^+(r)$, successeur par la ligne ℓ de l'arrêt multiple r.

On modifie aussi la définition des notations suivantes :

$\gamma^-(r)$, prédécesseur de l'arrêt simple r;

$\gamma^+(r)$, successeur de l'arrêt simple r;

R', ensemble des arrêts simples terminus de départ;

R", ensemble des arrêts simples terminus d'arrivée;

$\bar{\Gamma}^-(r) = \Gamma^-(r) - \gamma^-(r)$, $\quad r \in R^s - R'$;

$\bar{\Gamma}^-(r) = \Gamma^-(r) - \{\gamma_\ell^-(r),\ \ell \in L_r^-\}$, $\quad r \in R^m$;

$\bar{\Gamma}^+(r) = \Gamma^+(r) - \gamma^+(r)$, $\quad r \in R^s - R''$;

$\bar{\Gamma}^+(r) = \Gamma^+(r) - \{\gamma_\ell^+(r),\ \ell \in L_r^+\}$, $\quad r \in R^m$.

On aura sans doute remarqué que pour les arrêts multiples, nous n'avons pas défini d'ensemble des terminus d'arrivée ou de départ; nous avons été forcés de procéder différemment (en définissant les ensembles L_r^+ et L_r^-) parce qu'un arrêt multiple donné peut être à la fois le terminus d'arrivée d'un certain groupe de lignes, le terminus de départ d'un autre groupe et un arrêt intermédiaire pour les lignes restantes.

Il faut aussi noter qu'il peut maintenant exister dans le réseau augmenté des paires de sommets reliés par plusieurs arcs distincts (des arrêts successifs sur une portion commune à deux ou plusieurs lignes).

6.3 Caractérisation du comportement des usagers aux arrêts multiples et des stratégies

Un usager qui se présente à un arrêt multiple a la possibilité, s'il le désire, de prendre l'une ou l'autre des lignes desservant cet arrêt. De manière générale, on peut supposer qu'un certain nombre de ces lignes paraîtront attrayantes à l'usager, alors qu'il ne voudra, en aucun

cas, utiliser les autres (parce qu'elles ne lui permettent pas de se diriger vers sa destination ou qu'elles sont surchargées, par exemple). Dans le modèle, nous supposerons que le choix d'un usager se limite effectivement à la sélection d'un sous-ensemble de lignes attrayantes, et qu'il montera à bord du premier autobus d'une de ces lignes, dans lequel il trouvera une place (assise ou debout), bien qu'il soit tout à fait clair qu'il existe dans la réalité des choix plus complexes (par exemple, des choix faisant entrer en ligne de compte le temps déjà passé à attendre, le passage d'autobus pleins des lignes attrayantes, la possibilité d'avoir une place assise, etc.). Dans la suite, nous dirons que le sous-ensemble de lignes attrayantes choisi par un usager attendant à un arrêt multiple constitue la décision de celui-ci à cet arrêt.

Avec la notation que nous avons définie à la section précédente, on peut exprimer l'ensemble des décisions possibles pour les usagers attendant à un arrêt multiple donné r comme étant égal à $P(L_r^+) - \{\emptyset\}$, c'est-à-dire à l'ensemble des parties non-vides de L_r^+; par la suite, cet ensemble sera appelé Δ_r.

Une fois à bord d'un autobus, le passager doit décider à chaque arrêt desservi s'il va descendre ou poursuivre sa route dans le même autobus; encore une fois, il a une décision à prendre. Quand il descend, il doit alors se diriger vers un arrêt ou vers un centroïde : autre décision encore. Si l'on observe le cheminement d'un usager sur le graphe augmenté du réseau, on peut remarquer que les points de décision correspondent aux sommets du graphe et que, sauf aux arrêts multiples, les décisions possibles à un sommet correspondent identiquement aux arcs qui sont incidents

extérieurement à celui-ci (c'est-à-dire les arcs qui "partent" de ce sommet); aux arrêts multiples, il faut distinguer deux cas possibles : celui des usagers qui font accès au réseau de transport en commun (la décision qu'ils ont à prendre correspond alors à un choix de sous-ensemble de lignes attrayantes comme nous l'avons expliqué auparavant) et celui des usagers qui sont déjà dans un autobus; pour ceux-ci, on a plutôt une série de décisions à prendre :

. continuer à bord du même autobus ou descendre,

. si on descend, aller vers un autre sommet du réseau augmenté ou faire une correspondance au même arrêt,

. si on fait une correspondance, choisir le sous-ensemble de lignes attrayantes.

Sur le réseau augmenté, il est quelque peu difficile de visualiser clairement une telle séquence de décisions, c'est pourquoi nous ferons en sorte au moment de l'éclatement des arrêts multiples (section 6.5) de nous assurer que chaque décision corresponde à un choix d'arc à un sommet du réseau éclaté.

Quoi qu'il en soit, il est possible de définir complètement le cheminement d'un usager de son origine vers sa destination par ses décisions, c'est donc cela qui constitue sa <u>stratégie</u>. De manière plus précise, on dira que la stratégie d'un usager est un ensemble cohérent de décisions à certains sommets du réseau augmenté permettant à celui-ci de se rendre de son origine vers sa destination. Il faut noter ici que l'on n'a pas besoin de spécifier une décision pour tous les sommets du graphe, car certains

d'entre eux ne seront jamais visités par l'usager; il faut cependant en spécifier une (ou plusieurs dans le cas de certains arrêts multiples) pour tous les sommets qui peuvent être atteints par l'usager (c'est ce que nous voulons dire quand nous parlons d'"ensemble cohérent de décisions").

Par rapport à cette définition de la notion de stratégie, la solution ARSO de l'exemple 3.1 s'exprime comme suit :

. attente à l'arrêt A :
  sous-ensemble de lignes attrayantes = {1,2};

. passage à l'arrêt B dans un autobus de la ligne 1 :
  descendre; correspondre; attendre {2,3};

. passage à l'arrêt B dans un autobus de la ligne 2 :
  continuer;

. passage à l'arrêt C dans un autobus de la ligne 2 :
  continuer;

. passage à l'arrêt D dans un autobus de la ligne 2 :
  descendre; fin du déplacement;

. passage à l'arrêt D dans un autobus de la ligne 3 :
  descendre; fin du déplacement.

La caractérisation que nous venons de faire du comportement des usagers et des stratégies est très proche de celle de Spiess (1983), cependant, elle en diffère sur le point suivant : Spiess définit les stratégies par rapport aux destinations et non par rapport aux paires origines-destinations comme nous l'avons fait, de plus, dans sa définition, il faut

spécifier une décision pour tous les sommets du réseau. Nous avons préféré la formulation par paires O-D, car elle met en lumière l'analogie stratégies-itinéraires.

6.4 Définition du sous-modèle d'arrêt multiple

A tout arrêt multiple, nous avons besoin de pouvoir calculer le temps d'attente des usagers qui y prennent l'autobus et déterminer la répartition de ceux-ci entre les lignes attrayantes pour eux. Pour répondre à ce besoin, nous allons définir un sous-modèle d'arrêt multiple qui agira comme une "boîte noire" dans le modèle global d'équilibre.

Il est clair que le facteur dominant pour l'attente et la répartition est le choix de lignes (décision) que font les usagers. D'autre part, il semble tout à fait raisonnable de considérer l'origine et la destination des usagers comme des variables négligeables dans le problème qui nous intéresse. En définitive, on peut donc dire que le sous-modèle d'arrêt multiple doit produire, pour chaque choix possible de lignes attrayantes, le temps d'attente moyen pour tous les usagers faisant ce choix et la répartition de ceux-ci (sous forme de distribution de probabilité).

Essayons maintenant d'identifier les "entrées" (paramètres) que l'on doit fournir au sous-modèle. Il est clair (par analogie avec notre étude des arrêts simples) que l'on doit connaître le nombre moyen d'usagers faisant chaque choix de lignes par unité de temps, le nombre moyen de places libres dans les autobus de chaque ligne desservant l'arrêt (ou encore, le nombre moyen de places occupées et le nombre total de places disponibles),

ainsi que les lois de probabilité régissant les intervalles entre deux autobus successifs de chaque ligne, le nombre de places occupées à bord des autobus et les arrivées des usagers à l'arrêt.

Avant de poursuivre, il convient de souligner que le temps d'attente moyen et la répartition entre les lignes des usagers faisant un choix de lignes donné ne dépendent pas uniquement du taux d'arrivée de ces usagers et des paramètres de service des lignes choisies, mais de l'ensemble des taux d'arrivée et des paramètres de service de toutes les lignes desservant l'arrêt, car il y a un phénomène d'interaction assez fort entre les usagers. Ainsi, par exemple, si deux lignes passent à un arrêt, une augmentation du taux d'arrivée des usagers prenant uniquement l'une de celles-ci provoquera un changement dans la répartition des usagers pour lesquels les deux lignes sont attrayantes et, en conséquence, une variation dans le temps d'attente des usagers prenant uniquement l'autre ligne.

Si on pose pour un arrêt multiple r donné

$\tilde{v}_r^{\delta}$: nombre moyen d'usagers faisant le choix de lignes δ par unité de temps ($\delta \in \Delta_r$),

$\bar{v}_r^{\ell}$: nombre moyen de places occupées dans les autobus de la ligne ℓ par unité de temps ($\ell \in L_r^- \cap L_r^+$),

$\underline{\tilde{v}}_r$: $(\tilde{v}_r^{\delta})_{\delta \in \Delta_r}$,

$\underline{\bar{v}}_r$: $(\bar{v}_r^{\ell})_{\ell \in L_r^- \cap L_r^+}$,

on peut exprimer la dépendance du temps d'attente et de la répartition pour une décision δ donnée par rapport aux choix de tous les usagers et à la charge de toutes les lignes par les notations suivantes :

$\phi_r^{\delta}(\bar{\underline{v}}_r, \tilde{\underline{v}}_r)$: temps d'attente pour les usagers faisant le choix δ,

$\underline{p}_r^{\delta}(\bar{\underline{v}}_r, \tilde{\underline{v}}_r)$: vecteur de répartition pour les usagers faisant le choix δ.

Pour ce qui est de $\underline{p}_r^{\delta}$, il convient de souligner que

$$p_{r,\ell}^{\delta}(\bar{\underline{v}}_r, \tilde{\underline{v}}_r) = \Pr\{\text{prendre la ligne } \ell \text{ si on fait le choix } \delta \text{ étant donnés } \bar{\underline{v}}_r \text{ et } \tilde{\underline{v}}_r\},\ \ell \in \delta,\ \delta \in A_r,$$

et que

$$\sum_{\ell \in \delta} p_{r,\ell}^{\delta}(\bar{\underline{v}}_r, \tilde{\underline{v}}_r) = 1 , \quad \forall\ \delta \in \Delta_r .$$

Le lecteur attentif aura sans doute remarqué que seules les lignes appartenant à $L_r^+ \cap L_r^-$ ont été considérées dans la définition de $\bar{\underline{v}}_r$; cela s'explique par les deux raisons suivantes :

a) les lignes appartenant à $L_r - L_r^+$ ont leur terminus d'arrivée à l'arrêt r et, conséquemment, on peut les ignorer totalement pour tout ce qui concerne les usagers qui prennent l'autobus à cet arrêt;

b) les lignes appartenant à $L_r - L_r^-$ ont leur terminus de départ à l'arrêt r, ce qui nous assure que les autobus de ces lignes sont toujours vides quand ils se présentent à l'arrêt.

L'effet de la fréquence des autobus, de la taille des autobus sur les différentes lignes et de la nature des distributions régissant les arrivées des usagers, les intervalles interautobus et le nombre de places occupées dans les autobus n'est pas souligné dans la notation choisie pour les temps d'attente et les répartitions, car, pour un problème donné,

ce sont des paramètres fixes, alors que $\bar{\underline{v}}_r$ et $\tilde{\underline{v}}_r$ varient avec les affectations des usagers, ce sont des valeurs qui nous sont fournies par le modèle de flot.

On peut schématiser le sous-modèle d'arrêt multiple par le diagramme suivant :

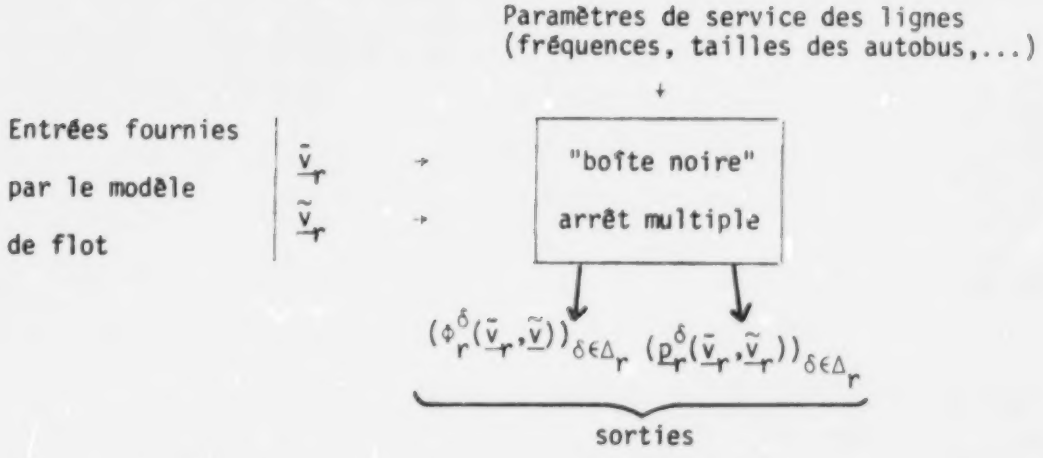


Figure 10 : Le sous-modèle d'arrêt multiple

Dans ce chapitre, nous nous contenterons d'utiliser les ϕ_r^δ et les p_r^δ dans le modèle d'équilibre sans indiquer comment on peut les déterminer, cela sera fait dans le chapitre suivant.

6.5 Eclatement des arrêts multiples

Comme dans le cas des arrêts simples (traité à la sous-sections 4.3.5), nous allons devoir éclater les arrêts multiples pour être en mesure de les représenter adéquatement. Dans ce processus d'éclatement, nous devons être attentifs aux choses suivantes :

a) la structure de lignes inhérente au réseau de transport en commun doit être mise en évidence (entre autres, en ce qui a trait aux correspondances);

b) dans le réseau éclaté, les décisions doivent correspondre à des choix d'arcs (cf. note de la section 6.3);

c) les entrées "variables" ($\underline{\bar{v}}_r$ et $\underline{\tilde{v}}_r$) du sous-modèle d'arrêt multiple doivent être faciles à calculer;

d) pour chaque choix de lignes possible, il faut un arc distinct d'attente dont le coût correspondra au temps d'attente moyen pour ce choix de lignes;

e) il faut une structure d'arcs et de sommets permettant de répartir les usagers qui prennent l'autobus selon les indications du sous-modèle d'arrêt multiple.

Examinons maintenant tour à tour chacune de ces exigences.

Pour maintenir l'intégrité de la structure de lignes, il est clair qu'il va falloir "séparer" les lignes de manière à avoir au moins un sommet pour chaque ligne desservant l'arrêt. On peut observer les effets de cette séparation sur la figure suivante (remarque : ici, on a laissé tomber les arcs de marche incidents à l'arrêt).

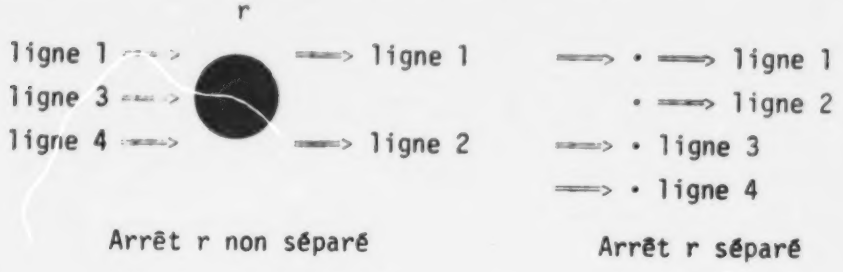


Figure 11 : Séparation des lignes

Une fois les lignes séparées, il est facile de voir que, pour déterminer le nombre moyen d'usagers à bord des autobus de chaque ligne ($\tilde{\underline{v}}_r$), on peut, comme dans le cas des arrêts simples, ajouter un arc auxiliaire pour chaque ligne pour laquelle l'arrêt qu'on éclate n'est pas un terminus. Si on reprend l'arrêt de la figure précédente, on obtient

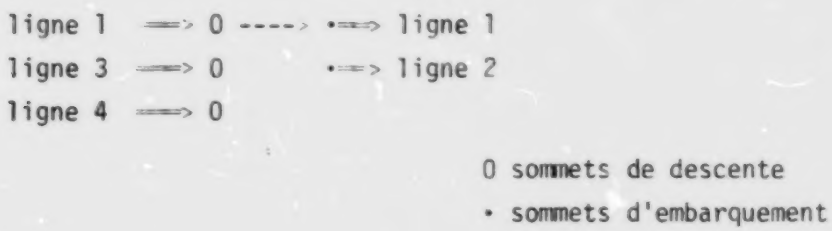


Figure 12 : Ajout des arcs auxiliaires

Pour déterminer le nombre moyen d'usagers faisant chaque choix de lignes ($\tilde{\underline{v}}_r$), il suffira de "ramasser" tous les usagers qui prennent l'autobus à un "sommet d'attente" et de faire "partir" de ce sommet un arc pour chaque choix de lignes. Le sommet terminal de chaque arc d'attente sera un "sommet de répartition" qui sera relié, par des "arcs de charge", aux sommets d'embarquement des lignes inclues dans le choix correspondant de manière à satisfaire la condition e).

Pour compléter l'éclatement, il ne reste plus qu'à ajouter un sommet auxiliaire dont le rôle principal sera de "ramasser" tous les usagers descendant à l'arrêt; en plaçant un "arc ce correspondance (immédiate)" de ce sommet vers le sommet d'attente, on fera en sorte que les usagers qui veulent effectuer des correspondances "passent" dans la structure d'attente-répartition.

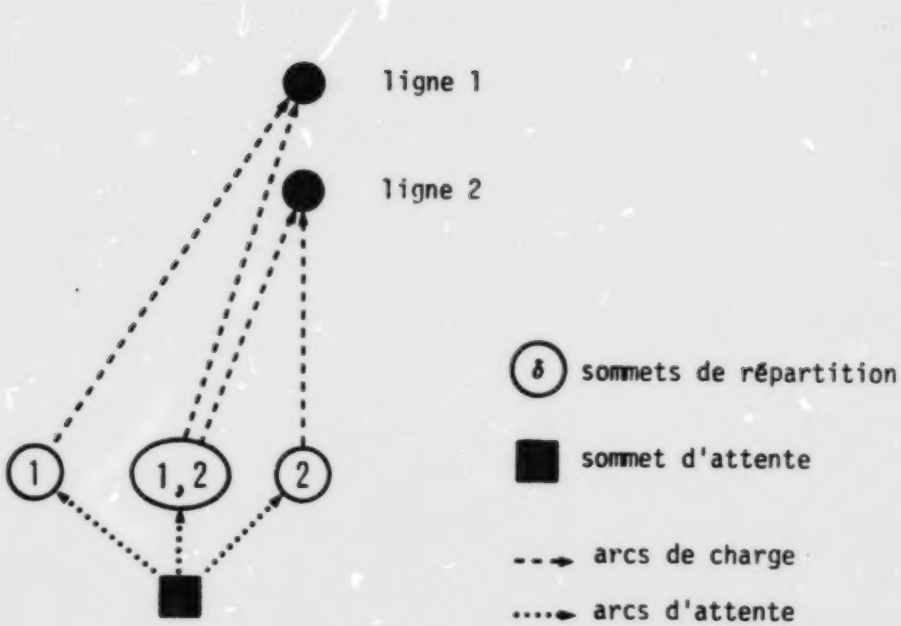


Figure 13 : Structure d'attente-répartition

Dans la représentation éclatée, les arcs de marche qui sont incidents extérieurement à l'arrêt seront incidents au sommet auxiliaire puisque c'est là que doivent "passer" tous les usagers qui descendent à l'arrêt, et ceux qui sont incidents intérieurement à l'arrêt seront incidents au sommet d'attente.

On peut maintenant vérifier que la condition b) que nous avons énoncée au départ est vérifiée :

. pour les usagers qui font accès au réseau, il n'y a qu'une seule décision à prendre, le choix de lignes qui correspond à un choix d'arc au sommet d'attente;

. pour les usagers à bord d'autobus, la décision "descendre ou continuer" correspond à choisir entre l'"arc de descente" (qui les emmène au sommet auxiliaire) ou l'arc auxiliaire quand ils se trouvent au sommet de descente de la ligne qu'ils utilisent;

. pour les usagers qui descendent, la décision "correspondre ou aller à un autre sommet du réseau augmenté" se ramène au choix d'arc au sommet auxiliaire;

. pour les usagers qui correspondent, le choix de lignes est fait comme pour ceux qui font accès au réseau.

En définitive, on obtient donc la représentation éclatée de la figure 15.

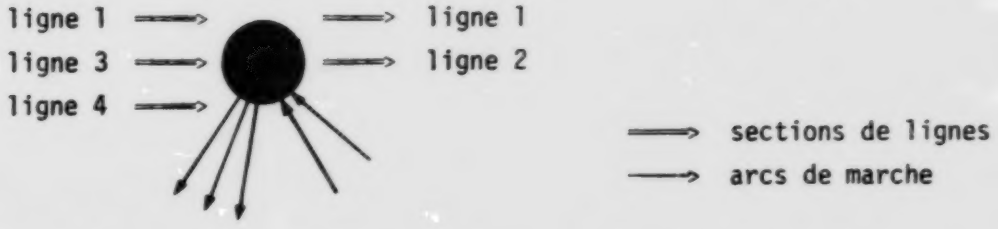


Figure 14 : L'arrêt multiple r dans le réseau augmenté.

Cette représentation éclatée n'est pas la plus économique en termes du nombre de sommets et d'arcs qu'il faut ajouter au réseau (on pourrait, entre autres choses, supprimer le sommet de répartition et l'arc de charge pour les choix qui se limitent à une ligne unique, en faisant

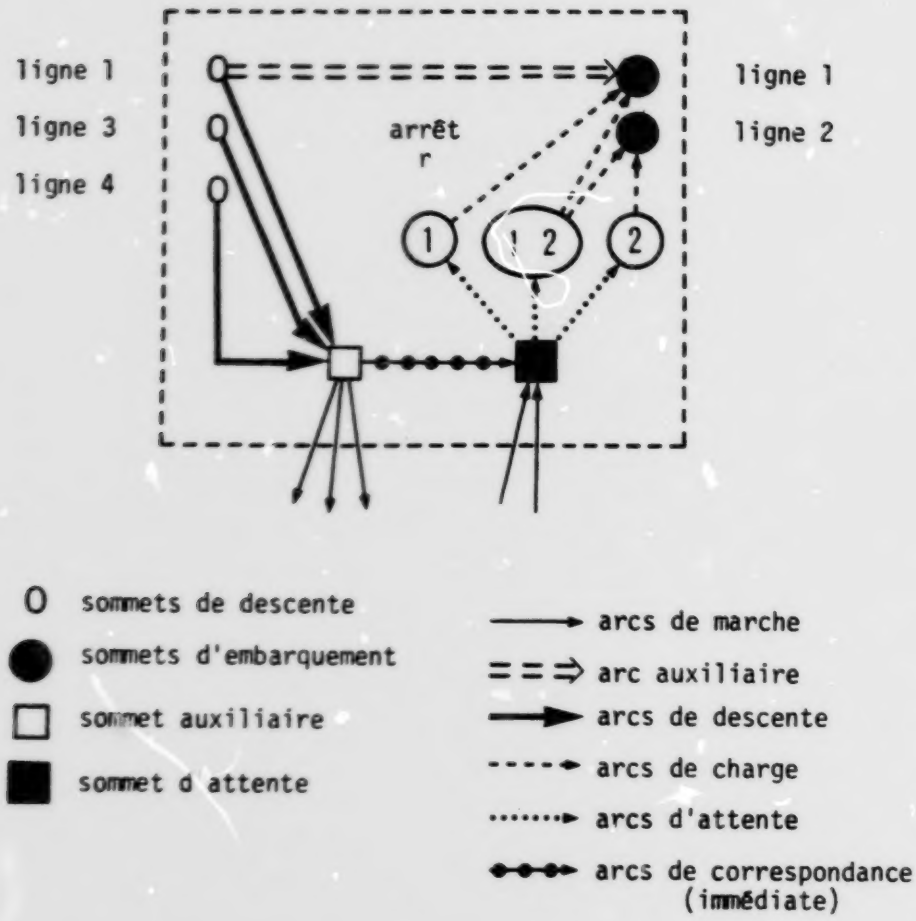


Figure 15 : L'arrêt r en représentation éclatée

aboutir l'arc d'attente au sommet d'embarquement de la ligne choisie), mais c'est celle qui est la plus facile à utiliser pour l'exposition du modèle; dans une implantation du modèle, on pourrait utiliser une représentation éclatée "concise" dans laquelle on ne garderait que les arcs et les sommets vraiment nécessaires.

Comme dans le chapitre quatrième, le réseau obtenu après éclatement des arrêts est appelé <u>réseau éclaté</u> et il est dénoté $[N^*,A^*]$.

Nous devons ajouter certaines notations et en redéfinir certaines autres pour le réseau éclaté. Les notations que l'on ajoute sont les suivantes :

r_ℓ^-, sommet de descente de la ligne ℓ à l'arrêt multiple r. $r \in R^m$, $\ell \in L_r^-$;

r_ℓ^+, sommet d'embarquement de la ligne ℓ à l'arrêt multiple r, $r \in R^m$, $\ell \in L_r^+$;

$b(r)$, sommet auxiliaire de l'arrêt multiple r, $r \in R^m$;

B, ensemble des sommets auxiliaires;

$e(r,\delta)$, sommet de répartition associé au choix de lignes δ à l'arrêt multiple r, $r \in R^m$, $\delta \in \Delta_r$;

E, ensemble des sommets de répartition;

E_r, ensemble des sommets de répartition de l'arrêt multiple r.

Nous redéfinissons aussi les notations suivantes :

r^-, sommet de descente de l'arrêt <u>simple</u> r, $r \in R^s - R'$;

r^+, sommet d'embarquement de l'arrêt <u>simple</u> r, $r \in R^s - R''$.

Avec cette notation, on peut identifier l'ensemble des arcs auxiliaires à l'intersection de A^* et du produit cartésien $R^- \times R^+$ (comme auparavant), l'ensemble des arcs d'attente des arrêts simples à l'intersection de A^* et du produit $U \times R^+$, l'ensemble des arcs d'attente des arrêts

multiples à l'intersection de A* et du produit U x E, l'ensemble des arcs de charge à l'intersection de A* et du produit E x R^+, l'ensemble des arcs de charge à l'intersection de A* et du produit R^- x B et l'ensemble des arcs de correspondance immédiate (qu'il faut distinguer des arcs de correspondance entre arrêts distincts contenus dans Q) à l'intersection de A* et du produit B x U.

Il faut noter ici que le temps de parcours des arcs de descente et de charge de la représentation éclatée est nul, tandis que celui des arcs d'attente est égal au temps d'attente du choix de lignes auquel ils correspondent, c'est-à-dire à $\Phi_r^{\delta}(\underline{\tilde{v}}_r, \underline{\tilde{v}}_r)$. Notons aussi que, si les unités sont définies de manière cohérente, $\tilde{v}_r^{\ell}$ est tout simplement le flot total sur l'arc auxiliaire de la ligne ℓ et $\tilde{v}_r^{\delta}$ le flot total sur l'arc d'attente pour le choix de lignes δ.

6.6 Variables de flot

Comme auparavant dans le chapitre quatrième, nous distinguerons dans le modèle les usagers en fonction de leur destination. D'autre part, la notation choisie pour les variables de flot sera définie autant que possible relativement aux sommets du réseau augmenté [N,A]. A cause de l'existence possible de plusieurs arcs distincts entre certains sommets du réseau augmenté (cf. remarque de la section 6.2), nous ajouterons, chaque fois qu'une ambiguité sera possible, un indice supplémentaire indiquant la ligne à laquelle appartient l'arc auquel nous voulons faire référence (la seule ambiguité possible consistant en effet à confondre des sections de lignes distinctes reliant les mêmes arrêts multiples). On posera donc :

v_{ij}^{d} ou v_{ij}^{ℓ}, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc (i,j), $(i,j)\in A$, l'indice ℓ pouvant être ajouté si $(i,j)\in S_{\ell}$.

Pour les arrêts simples, on aura encore :

$\bar{v}_{r}^{d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc auxiliaire de l'arrêt simple r;

$\tilde{v}_{r}^{d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc d'attente de l'arrêt simple r.

Pour les arrêts multiples, on aura :

$\bar{v}_{r}^{\ell d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc auxiliaire de la ligne ℓ à l'arrêt multiple r;

$\tilde{v}_{r}^{\delta d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc d'attente de la décision δ à l'arrêt multiple r;

$v_{r,\ell}^{+\delta d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc de charge de la ligne ℓ pour la décision δ à l'arrêt multiple r;

$v_{r,\ell}^{-d}$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc de descente de la ligne ℓ à l'arrêt multiple r;

$\hat{v}_r^d$, nombre moyen (par unité de temps) d'usagers se dirigeant vers la destination d qui empruntent l'arc de correspondance (immédiate) de l'arrêt multiple r.

Comme auparavant, on pourra obtenir facilement les flots totaux sur les arcs à partir de ces variables :

. $v_{ij} = \sum_{d \in D} v_{ij}^d$, $(i,j) \in A$

. $v_{ij}^\ell = \sum_{d \in D} v_{ij}^{\ell d}$, $(i,j) \in A$, $(i,j) \in S_\ell$

. $\bar{v}_r = \sum_{d \in D} \bar{v}_r^d$, $r \in R^s - R'$

. $\tilde{v}_r = \sum_{d \in D} \tilde{v}_r^d$, $r \in R^s - R''$

. $\bar{v}_r^\ell = \sum_{d \in D} \bar{v}_r^{\ell d}$. $r \in R^m$, $\ell \in L_r^- \cap L_r^+$

. $\tilde{v}_r^\delta = \sum_{d \in D} \tilde{v}_r^{\delta d}$, $r \in R^m$, $\delta \in \Delta_r$

. $v_{r,\ell}^{+\delta} = \sum_{d \in D} v_{r,\ell}^{+\delta d}$, $r \in R^m$, $\delta \in \Delta_r$, $\ell \in \delta$

. $v_{r,\ell}^- = \sum_{d \in D} v_{r,\ell}^{-d}$, $r \in R^m$, $\ell \in L_r^-$

. $\hat{v}_r = \sum_{d \in D} \hat{v}_r^d$, $r \in R^m$

Un vecteur de flots totaux sur le réseau éclaté sera encore appelé $\underline{v}$ ($\underline{v} = (v_a)_{a \in A^*}$).

6.7 Fonctions de coût sur les arcs

Pour les fonctions de coût sur les arcs, nous allons continuer d'utiliser la notation définie à la sous-section 4.3.8 dans laquelle, pour tout arc $a \in A^*$, on appelle $c_a(\underline{v})$ le coût induit par un vecteur de flots totaux $\underline{v}$ pour l'arc a. On pose aussi comme auparavant $\underline{c}(\underline{v}) = (c_a(\underline{v}))_{a \in A^*}$.

Si l'on fait la synthèse des résultats énoncés à la sous-section 4.3.8 et de ceux énoncés aux sections 6.4 et 6.5 relativement aux arrêts multiples, on peut exprimer $c_a(\underline{v})$ pour les différents types d'arcs comme suit :

. arcs de marche directe et d'accès

$$c_a(\underline{v}) = \alpha_m T_a, \qquad a \in M' \cup M''$$

. arcs de correspondance (entre arrêts distincts)

$$c_a(\underline{v}) = \alpha_c + \alpha_m T_a, \qquad a \in Q$$

. arcs de correspondance immédiate (arrêts multiples)

$$c_a(\underline{v}) = \alpha_c, \qquad a \in B \times U \ (\cap A^*)$$

. sections de lignes

$$c_a(\underline{v}) = T_a \Psi(v_a / K_\ell^0), \qquad a \in S_\ell, \ \ell \in L$$

. arcs auxiliaires (arrêts simples et multiples, arcs de descente (arrêts multiples) et arcs de charge (arrêts multiples)

$$c_a(\underline{v}) = 0, \qquad a \in R^- \times R^+ \ (\cap A^*)$$

$$a \in R^- \times B \ (\cap A^*)$$

$$a \in E \times R^+ \ (\cap A^*)$$

. arcs d'attente des arrêts simples

$$c_a(\underline{v}) = \alpha_a \Phi_r (\tilde{v}_r, \tilde{v}_r), \quad a = (u(r), r^+), \; r \in R^s - (R' \cup R'')$$

$$c_a(\underline{v}) = \alpha_a \Phi_r (0, \tilde{v}_r), \quad a = (u(r), r^+), \; r \in R'$$

. arcs d'attente des arrêts multiples

$$c_a(\underline{v}) = \alpha_a \Phi_r^{\delta}(\underline{\tilde{v}}_r, \underline{\tilde{v}}_r), \quad a = (u(r), e(r,\delta)), \; \delta \in \Delta_r, \; r \in R^m .$$

6.8 Affectations réalisables

L'ajout des arrêts multiples au réseau de transport en commun ne modifie pas énormément les conditions que nous avons énoncées à la section 4.4 pour qu'une affectation soit réalisable : satisfaction des demandes de transport et conservation du flot aux divers sommets du réseau. Il faut néanmoins remarquer qu'aux sommets de répartition des arrêts multiples, on n'a pas des conditions ordinaires de conservation du flot, mais plutôt des conditions de répartition des usagers, car ceux-ci doivent être répartis entre les lignes selon la distribution de probabilité qui est spécifiée par le sous-modèle d'arrêt multiple. Il convient de s'attarder quelque peu sur ce qui se passe à ces arrêts.

Considérons le cas du sommet de répartition associé à la décision δ d'un arrêt multiple quelconque r. Dans la section 6.4, on a fait l'hypothèse que l'origine et la destination des usagers n'étaient pas des variables pertinentes au niveau du sous-modèle d'arrêt multiple. Cela revient en quelque sorte à dire qu'il faut, pour une décision donnée, répartir les usagers se dirigeant vers toutes les destinations de la même manière dans le modèle de flot. Soit $\tilde{v}_r^{\delta d}$ le nombre moyen d'usagers se dirigeant vers

une destination d et prenant la décision δ et soit ℓ une ligne donnée appartenant à δ, le nombre moyen d'usagers se dirigeant vers d qui prendront la ligne ℓ suite à la décision $(v_{r,\ell}^{+\delta d})$ est obtenu en multipliant $v_r^{\delta d}$ par la probabilité pour chacun d'entre eux de prendre la ligne ℓ, autrement dit par $p_{r,\ell}^{\delta}(\underline{\bar{v}}_r, \underline{\tilde{v}}_r)$ (remarque : ici, nous travaillons en fait en termes d'espérance mathématique du nombre d'usagers qui prendront la ligne ℓ). C'est donc dire que l'on doit avoir

$$v_{r,\ell}^{+\delta d} = p_{r,\ell}^{\delta}(\underline{\bar{v}}_r, \underline{\tilde{v}}_r)\ \tilde{v}_r^{\delta d} \quad .$$

Avant de poursuivre, soulignons que ces conditions de répartition établissent un lien quelque peu étrange et pour le moins inhabituel dans un modèle de flot entre certaines des variables. Si l'on considère de plus le fait que, a priori, les $p_r^{\delta}\ (\underline{\bar{v}}_r, \underline{\tilde{v}}_r)$ ne sont pas définies sous forme d'expressions explicites des $\underline{v}_r$ et $\underline{\tilde{v}}_r$, il devient relativement évident que la nature même du domaine des affectations réalisables reste à clarifier.

Quoi qu'il en soit, nous allons écrire complètement les conditions que doivent satisfaire les affectations réalisables en utilisant les deux systèmes de notation (réseau éclaté, réseau augmenté) que nous avons définis.

Avec la notation définie pour le réseau éclaté, on obtient :

$$(1') \qquad \sum_{j \in \tilde{\Gamma}^+(d)} v_{dj}^d - \sum_{j \in \tilde{\Gamma}^-(d)} v_{jd}^d = -\sum_{o \in O} g^{od}, \quad d \in D$$

$$(2') \quad \sum_{j\in\tilde{\Gamma}^+(o)} v^d_{oj} - \sum_{j\in\tilde{\Gamma}^-(o)} v^d_{jo} = g^{od}, \quad d\in D,\ o\in O-\{d\}$$

$$(3') \quad \sum_{j\in\tilde{\Gamma}^+(i)} v^d_{ij} - \sum_{j\in\tilde{\Gamma}^-(i)} v^d_{ji} = 0, \quad d\in D,\ i\in N^*-(O\cup\{d\}\cup E)$$

$$(4') \quad v^{+\delta d}_{r,\ell} - p^{\delta}_{r,\ell}(\underline{\tilde{v}}_r, \underline{\tilde{v}}_r)\tilde{v}^{\delta d}_r = 0, \quad d\in D,\ r\in R^m,\ \delta\in\Delta_r,\ \ell\in\delta$$

$$(5') \quad v^d_{ij} \geq 0, \quad d\in D,\ (i,j)\in A^*$$

Les équations (1') à (3') sont identiques à celles que nous avions dans le modèle sans arrêts multiples; les équations (4') sont tout simplement les équations de répartition réécrites en ramenant tous les termes à gauche du signe d'égalité; les inéquations (5') sont les mêmes que dans l'autre modèle.

Avec la notation définie pour le réseau augmenté, on obtient, en distinguant les types de sommets :

$$(1) \quad \sum_{j\in\Gamma^+(d)} v^d_{dj} - \sum_{j\in\Gamma^-(d)} v^d_{jd} = \sum_{o\in O} g^{od}, \quad d\in D$$

$$(2) \quad \sum_{j\in\Gamma^+(o)} v^d_{oj} - \sum_{j\in\Gamma^-(o)} v^d_{jo} = g^{od}, \quad d\in D,\ o\in O-\{d\}$$

$$(3) \quad \sum_{j\in\Gamma^+(z)} v^d_{zj} - \sum_{j\in\Gamma^-(z)} v^d_{jz} = 0, \quad d\in D,\ z\in Z-(O\cup\{d\})$$

$$(4) \quad \sum_{j\in\bar{\Gamma}^+(r)} v^d_{rj} + \bar{v}^d_r - v^d_{\gamma^-(r),r} = 0, \quad d\in D,\ r\in R^s-(R'\cup R'')$$

$$(5) \quad \tilde{v}_r^d - \sum_{j \in \Gamma^-(r)} v_{jr}^d = 0, \quad d \in D, \; r \in R^s - (R' \cup R'')$$

$$(6) \quad v_{r,\gamma^+(r)}^d - \bar{v}_r^d - \tilde{v}_r^d = 0, \quad d \in D, \; r \in R^s - (R' \cup R'')$$

$$(7) \quad \tilde{v}_r^d - \sum_{j \in \Gamma^-(r)} v_{jr}^d = 0, \quad d \in D, \; r \in R'$$

$$(8) \quad v_{r,\gamma^+(r)}^d - \tilde{v}_r^d = 0, \quad d \in D, \; r \in R'$$

$$(9) \quad \sum_{j \in \Gamma^+(r)} v_{rj}^d - v_{\gamma^-(r),r}^d = 0, \quad d \in D, \; r \in R''$$

$$(10) \quad \bar{v}_r^{\ell d} + v_{r,\ell}^{-d} - v_{\gamma_\ell^-(r),r}^{\ell d} = 0, \quad d \in D, \; r \in R^m, \; \ell \in L_r^- \cap L_r^+$$

$$(11) \quad v_{r,\ell}^{-d} - v_{\gamma_\ell^-(r),r}^{\ell d} = 0, \quad d \in D, \; r \in R^m, \; \ell \in L_r^- - L_r^+$$

$$(12) \quad \sum_{j \in \Gamma^+(r)} v_{rj}^d + \hat{v}_r^d - \sum_{\ell \in L_r^-} v_{r,\ell}^{-d} = 0, \quad d \in D, \; r \in R^m$$

$$(13) \quad \sum_{\delta \in \Delta_r} \tilde{v}_r^{\delta d} - \hat{v}_r^d - \sum_{j \in \Gamma^-(r)} v_{jr}^d = 0 \quad d \in D, \; r \in R^m$$

$$(14) \quad v_{r,\gamma_\ell^+(r)}^{\ell d} - \sum_{\delta \mid \ell \in \delta} v_{r,\ell}^{+\delta d} - \bar{v}_r^{\ell d} = 0, \quad d \in D, \; r \in R^m, \; \ell \in L_r^+ \cap L_r^-$$

$$(15) \quad v_{r,\gamma_\ell^+(r)}^{\ell d} - \sum_{\delta \mid \ell \in \delta} v_{r,\ell}^{+\delta d} = 0, \quad d \in D, \; r \in R^m, \; \ell \in L_r^+ - L_r^-$$

$$(16) \quad v_{r,\ell}^{+\delta d} - p_{r,\ell}^{\delta}(\underline{\tilde{v}}_r, \underline{\tilde{v}}_r)\tilde{v}_r^{\delta d} = 0, \quad d \in D, \; r \in R^m, \; \delta \in \Delta_r, \; \ell \in \delta$$

(17) Non-négativité de toutes les variables.

Les équations (1) à (9) sont celles-là mêmes que nous avions dans le modèle sans arrêts multiples (sauf qu'on a remplacé R par R^S).

Les contraintes (10) et (11) sont les équations de conservation aux sommets de descente des arrêts multiples (il faut séparer le cas où l'arrêt r est le terminus d'arrivée de la ligne ℓ de celui où il ne l'est pas).

Les contraintes (12) et (13) sont les équations de conservation de flot aux sommets auxiliaires et aux sommets d'attente des arrêts multiples.

Les contraintes (14) et (15) sont les équations de conservation de flot aux sommets d'embarquement des arrêts multiples (en distinguant les terminus).

Les équations (16) sont les équations de répartition aux arrêts multiples (identiques à (5') dans la notation précédente).

6.9 Affectations d'équilibre

Voyons maintenant à quelles conditions une affectation réalisable (au sens de la section précédente) constitue une affectation d'équilibre des passagers. Pour ce faire, nous pouvons procéder comme dans le chapitre quatrième (section 4.5).

Supposons que nous ayons en main une affectation réalisable $\{v_a^{d*}\}_{a\in A^*}^{d\in D}$. Soit $\underline{v}^*$ le vecteur des flots totaux correspondant à cette affectation, $\underline{c}(\underline{v}^*)$ le vecteur de coûts induit par v^* et $\{p_r^{\delta}(\bar{\underline{v}}_r^*, \tilde{\underline{v}}^*)\}_{r\in R^m}^{\delta\in\Delta_r}$

l'ensemble des vecteurs de répartition induits par $\underline{v}^*$. Il s'agit maintenant de déterminer si, en se basant sur ces coûts et ces répartitions, certains usagers peuvent diminuer l'espérance de leur coût de transport en changeant de stratégie. On peut remarquer ici que, une fois fixés les coûts et les répartitions, le problème de recherche d'une stratégie à espérance de coût minimale pour un couple de sommets du réseau correspond à un programme dynamique stochastique. Heureusement, il n'est pas nécessaire de résoudre un programme dynamique pour chaque paire O-D. On peut en effet reformuler légèrement la question de manière à obtenir par un seul programme les stratégies à espérance de coût minimale de tous les sommets du réseau vers une destination d donnée. Si on pose

$$\Pi_i^d(\underline{v}^*) = E\,[\text{coût de la stratégie optimale pour aller de } i \text{ à } d \mid \underline{v}^*\,], \quad i\in N^*,\ d\in D,$$

on peut écrire les relations suivantes :

$$\Pi_d^d(\underline{v}^*) = 0,\ d\in D \quad \text{(par définition)},$$

$$\Pi_i^d(\underline{v}^*) = \min_{j\in\tilde{\Gamma}^+(i)} \{c_{ij}(\underline{v}^*) + \Pi_j^d(\underline{v}^*)\},\quad i\in N^*-(E\cup\{d\}),\ d\in D,$$

$$\Pi_i^d(\underline{v}^*) = \sum_{\ell\in\delta} p^{\delta}_{r,\ell}\,(\bar{\underline{v}}^*_r,\tilde{\underline{v}}^*_r)\,\Pi^d_{r^+_\ell}(\underline{v}^*),\quad i=e(r,\delta),\ r\in R^m,\ \delta\in\Delta_r.$$

On peut remarquer que pour les sommets de N*-E, la relation définissant Π_i^d est semblable à celle que l'on obtient dans un calcul de chemin le plus court (cf. sous-section 4.5.2), car, à ces sommets, on peut établir une équivalence <choix de décision> <=> <choix d'arc>. Pour les sommets de répartition (ceux de E), on n'a pas à calculer de minimum, car

l'usager ne contrôle pas son choix d'arc; plus précisément, aux sommets de répartition, l'expression définissant Π_i^d correspond simplement à un calcul d'espérance mathématique pour le coût de la suite du parcours.

Si le réseau éclaté est fortement connexe par rapport aux destinations, on aura toujours

$$\Pi_i^d (\underline{v}^*) < +\infty , \qquad \forall\, d\in D,\ \forall\, i\in N^*.$$

Dans un tel cas, on peut affirmer que pour $i\in N^*-E$ et $j\in\tilde{\Gamma}^+(i)$, l'égalité

$$c_{ij} (\underline{v}^*) = \Pi_i^d (\underline{v}^*) - \Pi_j^d(\underline{v}^*)$$

implique l'optimalité de la "décision (i,j)" pour le sommet i.

Faisant abstraction des arcs de charge pour un moment, on peut donc construire pour chaque destination un <u>sous-graphe des stratégies optimales</u> vers cette destination en utilisant la règle suivante :

$$\langle (i,j)\in\{\text{sous-graphe des stratégies optimales vers d}\}\rangle$$

$$\Updownarrow$$

$$\langle i\in N^*-E \text{ et } c_{ij}(\underline{v}^*) = \Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*)\rangle .$$

Revenant maintenant à la question des arcs de charge, on peut remarquer qu'ils appartiennent tous d'une certaine façon aux sous-graphes des stratégies optimales puisque, pour aller d'un sommet de répartition à n'importe quel autre sommet du réseau, il faut "utiliser" les arcs de charge incidents aux sommets de répartition dans les proportions définies par $\{p_r^{\delta}(\tilde{\underline{v}}_r^*, \tilde{\underline{v}}_r^*)\}$.

Tout ceci ayant été posé, il devient facile de voir qu'un usager peut réduire l'espérance de son coût de déplacement si, et seulement si, sa stratégie contient des arcs n'appartenant pas au sous-graphe des stratégies optimales vers sa destination.

Si on appelle $\bar{A}^*$ l'ensemble des arcs du réseau éclaté qui ne sont pas des arcs de charge ($\bar{A}^* = A^* - E \times R^+$), on peut formuler la condition d'équilibre comme suit.

Théorème 6.1 : Condition d'équilibre

Soit $\{v_a^{d*}\}_{a\in A^*}^{d\in D}$ une affectation réalisable des usagers, $\underline{v}^*$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v}^*)$ le vecteur de coûts induit par $\underline{v}^*$, $\{\underline{p}_r^{\delta}(\underline{\tilde{v}}_r^*, \underline{\tilde{v}}_r^*)\}_{r\in R^m}^{\delta\in\Delta_r}$ l'ensemble des vecteurs de répartition induits par v^* et $(\Pi_i^d(\underline{v}^*))_{i\in N^*}^{d\in D}$ le vecteur des espérances de coût des stratégies optimales pour $\underline{c}(\underline{v}^*)$ et $\{\underline{p}_r^{\delta}(\underline{\tilde{v}}_r^*, \underline{\tilde{v}}_r^*)\}$, $\{v_a^{d*}\}$ est une affectation d'équilibre si, et seulement si,

$$\forall\ (i,j)\in\bar{A}^*,\quad \forall\ d\in D,\ v_{ij}^{d*}(\Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*) - c_{ij}(\underline{v}^*)) = 0\ .$$

Démonstration

Suite à la discussion qui précède, on voit facilement que

<$\{v_a^{d*}\}$ est une affectation d'équilibre>

<=> <$\forall\ d\in D$, les seuls arcs utilisés par les usagers se dirigeant vers d appartiennent au sous-graphe des stratégies optimales vers d>

<=> <$\forall\ d\in D$, $\forall (i,j)\in A^*$. $v_{ij}^{d*} > 0$ => (i,j) appartient au sous-graphe des stratégies optimales vers d> .

Pour $(i,j) \in A^* - \bar{A}^*$, on a toujours l'appartenance de (i,j) au sous-graphe des stratégies optimales vers toutes les destinations. La dernière condition est donc équivalente à

$$\langle \forall\, d \in D,\ \forall\, (i,j) \in \bar{A}^*,\ v_{ij}^{d*} > 0 \Rightarrow (i,j) \text{ appartient au sous-graphe des stratégies optimales vers } d \rangle$$

$$\Leftrightarrow \quad \langle \forall\, d \in D,\ \forall\, (i,j) \in \bar{A}^*,\ v_{ij}^{d*} > 0 \Rightarrow c_{ij}(\underline{v}^*) = \Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*) \rangle\ .$$

Remarquons maintenant que, pout toute affectation réalisable $\{v_a^{d*}\}$, pour toute destination d et pour tout arc $(i,j) \in \bar{A}^*$, om a $v_{ij}^{d*} \geq 0$ et qu'alors, les conditions

$$\langle v_{ij}^{d*} > 0 \Rightarrow c_{ij}(\underline{v}^*) = \Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*) \rangle$$

et

$$\langle v_{ij}^{d*} \left(\Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*) - c_{ij}(\underline{v}^*)\right) = 0 \rangle$$

sont équivalentes, ce qui nous permet de conclure que

$$\langle \{v_a^{d*}\} \text{ est une affectation d'équilibre} \rangle$$

$$\Leftrightarrow \quad \langle \forall\, (i,j) \in \bar{A}^*,\ \forall\, d \in D,\ v_{ij}^{d*} \left\{\Pi_i^d(\underline{v}^*) - \Pi_j^d(\underline{v}^*) - c_{ij}(\underline{v}^*)\right) = 0 \rangle$$

□

On peut maintenant définir une mesure de déséquilibre comme pour le modèle sans arrêts multiples (cf. sous-section 4.5.4). La seule différence notable entre les deux définitions est la substitution de $\bar{A}^*$ pour A^* dans celle qui suit.

Définition 6.1 : Soit $\{v_a^d\}_{a\in A^*}^{d\in D}$ une affectation réalisable quelconque, $\underline{v}$ le vecteur des flots totaux qui y correspond, $c(\underline{v})$ le vecteur de coûts induit par $\underline{v}$, $\{\underline{p}_r^{\delta}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)\}_{r\in R^m}^{\delta\in\Delta_r}$ l'ensemble des vecteurs de répartition induits par $\underline{v}$ et $(\Pi_i^d(\underline{v}))_{i\in N^*}^{d\in D}$ le vecteur des espérances de coût des stratégies optimales pour $\underline{c}(\underline{v})$ et $\{p_r^{\delta}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)\}$, on appelle mesure de déséquilibre de $\{v_a^d\}$ la quantité $\tau\{v_a^d\}$ suivante :

$$\tau\{v_a^d\} = \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d\,(c_{ij}(\underline{v})-(\Pi_i^d(\underline{v})-\Pi_j^d(\underline{v}))).$$

Il est clair qu'ici encore, la mesure de déséquilibre de toute affectation réalisable est non-négative et que les affectations d'équilibre (et seulement elles) ont une mesure de déséquilibre nulle (propriétés 4.1 et 4.2). On peut aussi démontrer un théorème "équivalent" au théorème 4.2 que nous avions pour l'autre version du modèle.

Théorème 6.2 : Soit $\{v_a^d\}_{a\in A^*}^{d\in D}$ une affectation réalisable, $\underline{v}$ le vecteur des flots totaux qui y correspond, $\underline{c}(\underline{v})$ le vecteur de coûts induit par v, $\{\underline{p}_r(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)\}_{r\in R^m}^{\delta\in\Delta_r}$ l'ensemble des vecteurs de répartition induits par $\underline{v}$ et $(\Pi_i^d(\underline{v}))_{i\in N^*}^{d\in D}$ le vecteur des espérances de coûts des stratégies optimales pour $\underline{c}(\underline{v})$ et $\{\underline{p}_r^{\delta}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)\}$, alors on a

$$\tau\{v_a^d\} = \sum_{(i,j)\in A^*} c_{ij}(\underline{v})v_{ij} - \sum_{d\in D}\sum_{o\in O}\Pi_o^d(\underline{v})\,g^{od}$$

et

$$\tau\{v_a^d\} = \sum_{(i,j)\in A^*} c_{ij}(\underline{v})v_{ij} - \sum_{d\in D}\sum_{o\in O}\Pi_o^d(\underline{v})\,g^{od}\ .$$

Démonstration

On a

$$\tau\{v_a^d\} = \sum_{d\in D} \sum_{(i,j)\in \bar{A}^*} v_{ij}^d \left(c_{ij}(\underline{v}) - (\Pi_i^d(\underline{v}) - \Pi_j^d(\underline{v}))\right)$$

$$\Rightarrow \quad \tau\{v_a^d\} = \sum_{d\in D} \sum_{(i,j)\in \bar{A}^*} v_{ij}^d\, c_{ij}(\underline{v}) - \sum_{d\in D} \sum_{(i,j)\in \bar{A}^*} v_{ij}^d \left(\Pi_i^d(\underline{v}) - \Pi_j^d(\underline{v})\right)$$

Calculons maintenant $\sum_{d\in D} \sum_{(i,j)\in A^*-\bar{A}^*} v_{ij}^d \left(\Pi_i^d(\underline{v}) - \Pi_j^d(\underline{v})\right)$.

On a $A^* - \bar{A}^* = A^* \cap (E \times R^+)$

$$= \bigcup_{r\in R^m} \left[A^* \cap (E_r \times R^+)\right]$$

$$= \bigcup_{r\in R^m} \bigcup_{\delta\in\Delta_r} \left[A^* \cap (\{e(r,\delta)\} \times R^+)\right]$$

$$A^* - \bar{A}^* = \bigcup_{r\in R^m} \bigcup_{\delta\in\Delta_r} \bigcup_{\ell\in\delta} \{(e(r,\delta), r_\ell^+)\}$$

On en tire

$$\sum_{d\in D} \sum_{(i,j)\in A^*-\bar{A}^*} v_{ij}^d \left(\Pi_i^d(\underline{v}) - \Pi_j^d(\underline{v})\right) =$$

$$\sum_{d\in D} \sum_{r\in R^m} \sum_{\delta\in\Delta_r} \sum_{\ell\in\delta} v_{r,\ell}^{+\delta d} \left(\Pi_{e(r,\delta)}^d(\underline{v}) - \Pi_{r_\ell^+}^d(\underline{v})\right)$$

(Remarque : $v_{r,\ell}^{+\delta d}$ est le flot se dirigeant vers d sur l'arc $(e(r,\delta), r_\ell^+)$)

Pour d, r et δ fixés, examinons la sommation

$$\sum_{\ell \in \delta} v_{r,\ell}^{+\delta d} \left(\Pi^{d}_{e(r,\delta)}(\underline{v}) - \Pi^{d}_{r_\ell^+}(\underline{v})\right).$$

On peut écrire

$$\sum_{\ell \in \delta} v_{r,\ell}^{+\delta d}\left(\Pi^{d}_{e(r,\delta)}(\underline{v}) - \Pi^{d}_{r_\ell^+}(\underline{v})\right) = \Pi^{d}_{e(r,\delta)} \sum_{\ell \in \delta} v_{r,\ell}^{+\delta d} - \sum_{\ell \in \delta} v_{r,\ell}^{+\delta d} \Pi^{d}_{r_\ell^+}(\underline{v})$$

Remarquons ici que $\sum_{\delta \in \ell} v_{r,\ell}^{+\delta d} = \tilde{v}_r^{\delta d}$

et que $v_{r,\ell}^{\delta d} = p^{\delta}_{r,\ell}\,(\tilde{\underline{v}}_r, \tilde{\underline{v}}_r)\, \tilde{v}_r^{\delta r}$, $\ell \in \delta$.

De ceci, on tire que

$$\Pi^{d}_{e(r,\delta)} \sum_{\ell \in \delta} v_{r,\ell}^{+\delta d} = \Pi^{d}_{e(r,\delta)} \tilde{v}_r^{\delta d}$$

et que

$$\sum_{\ell \in \delta} v_{r,\ell}^{+\delta d}\, \Pi^{d}_{r_\ell^+}(\underline{v}) = \sum_{\ell \in \delta} p^{\delta}_{r,\ell}(\tilde{\underline{v}}_r, \tilde{\underline{v}}_r) \tilde{v}_r^{\delta d}\, \Pi^{d}_{r_\ell^+}(\underline{v})$$

$$= \tilde{v}_r^{\delta d} \sum_{\ell \in \delta} p^{\delta}_{r,\ell}(\tilde{\underline{v}}_r, \tilde{\underline{v}}_r)\, \Pi^{d}_{r_\ell^+}(\underline{v})$$

$$\Rightarrow \quad \sum_{\ell \in \delta} r_{r,\ell}^{+\delta d}\, \Pi^{d}_{r_\ell^+}(\underline{v}) = \tilde{v}_r^{\delta d}\, \Pi^{d}_{e(r,\delta)}(\underline{v}) .$$

On peut donc conclure que

$$\sum_{\ell \in \delta} v_{r,\ell}^{+\delta d}\left(\Pi^{d}_{e(r,\delta)}(\underline{v}) - \Pi^{d}_{r_\ell^+}(\underline{v})\right) = 0, \ \forall\, d \in D, \ \forall\, r \in R^m, \ \forall \delta \in \Delta_r$$

et donc

$$\sum_{d \in D} \sum_{(i,j) \in A^* - \bar{A}^*} v_{ij}^{d}\left(\Pi_i^d(\underline{v}) - \Pi_j^d(\underline{v})\right) = 0$$

Ceci nous permet de réécrire $\tau\{v_a^d\}$ sous la forme

$$\tau\{v_a^d\} = \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d c_{ij}(\underline{v}) - \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d (\pi_i^d(\underline{v}) - \pi_j^d(\underline{v}))$$

ou encore

$$\tau\{v_a^d\} = \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d c_{ij}(\underline{v}) - \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d \pi_i^d(\underline{v}) + \sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d \pi_j^d(\underline{v}) \qquad (1)$$

Examinons maintenant chacune des doubles sommations séparément. On a

$$\sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d c_{ij}(\underline{v}) = \sum_{(i,j)\in A^*} c_{ij}(\underline{v}) \sum_{d\in D} v_{ij}^d = \sum_{(i,j)\in A^*} c_{ij}(\underline{v}) v_{ij} \qquad (2)$$

D'autre part,

$$\sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d \pi_i^d(\underline{v}) = \sum_{d\in D} \sum_{i\in N^*} \pi_i^d(\underline{v}) \sum_{j\in \tilde{\Gamma}^+(i)} v_{ij}^d \qquad (3)$$

et

$$\sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d \pi_j^d(\underline{v}) = \sum_{d\in D} \sum_{j\in N^*} \pi_j^d(\underline{v}) \sum_{i\in \tilde{\Gamma}^-(j)} v_{ij}^d \qquad (4)$$

par les identités

$$A^* = \bigcup_{i\in N^*} \{(i,j) | j\in \tilde{\Gamma}^+(i)\} = \bigcup_{j\in N^*} \{(i,j) | i\in \tilde{\Gamma}^-(j)\}.$$

En interchangeant le rôle des indices dans le membre droit de (4), on obtient

$$\sum_{d\in D} \sum_{(i,j)\in A^*} v_{ij}^d \pi_j^d(\underline{v}) = \sum_{d\in D} \sum_{i\in N^*} \pi_i^d(\underline{v}) \sum_{j\in \tilde{\Gamma}^-(i)} v_{ji}^d \qquad (5)$$

En combinant (3) et (5), on peut tirer

$$\sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \Pi_j^d(\underline{v}) - \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \Pi_i^d(\underline{v}) =$$

$$\sum_{d\in D}\sum_{i\in N^*} \Pi_i^d(\underline{v}) \sum_{j\in\tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{d\in D}\sum_{i\in N^*} \Pi_i^d(\underline{v}) \sum_{j\in\tilde{\Gamma}^+(i)} v_{ij}^d =$$

$$\sum_{d\in D}\sum_{i\in N^*} \Pi_i^d(\underline{v}) \left[\sum_{j\in\tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j\in\tilde{\Gamma}^+(i)} v_{ij}^d \right] \qquad (6)$$

Pour d fixé, on a les résultats suivants :

$$\sum_{j\in\tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j\in\tilde{\Gamma}^+(i)} v_{ij}^d = 0, \qquad \forall\, i\in N^* - (0\cup\{d\})$$

(contraintes (3') et (4') de système définissant les affectations réalisables)

$$\sum_{j\in\tilde{\Gamma}^-(i)} v_{ji}^d - \sum_{j\in\tilde{\Gamma}^+(i)} v_{ij}^d = -g^{id}, \qquad \forall\, i\in(0-\{d\})$$

$$\Pi_d^d(\underline{v}) = 0 \qquad \text{(définition de } \Pi\text{)}$$

L'équation (6) est donc équivalente à

$$\sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \Pi_j^d(\underline{v}) - \sum_{d\in D}\sum_{(i,j)\in A^*} v_{ij}^d \Pi_i^d(\underline{v}) = -\sum_{d\in D}\sum_{o\in O} \Pi_o^d(\underline{v}) g^{od} \qquad (7)$$

En substituant (2) et (7) dans (1), on obtient

$$\tau\{v_a^d\} = \sum_{(i,j)\in\bar{A}^*} c_{ij}(\underline{v}) v_{ij} - \sum_{d\in D}\sum_{o\in O} \Pi_o^d(\underline{v}) g^{od}$$

$$\Leftrightarrow \tau\{v_a^d\} = \sum_{(i,j)\in A^*} c_{ij}(\underline{v}) v_{ij} - \sum_{d\in D}\sum_{o\in O} \Pi_o^d(\underline{v}) g^{od}$$

(étant donné que $c_{ij}(\underline{v}) = 0, \quad \forall\, (i,j)\in A^*-\bar{A}^*$)

□

Ce théorème nous permet de caractériser la notion d'équilibre par rapport aux vecteurs de flots totaux comme dans le chapitre quatrième : tous les résultats de la sous-section 4.5.5 s'appliquent donc également au modèle d'équilibre pour les réseaux comportant des arrêts multiples.

6.10 Quelques remarques

Suite aux résultats de la section précédente, on peut être amené à penser que la présence d'arrêts multiples dans un réseau de transport ne complique pas énormément le problème de l'affectation des passagers; cependant, il faut souligner les faits suivants :

a) Alors que pour éclater un arrêt simple on a besoin de trois sommets et de deux arcs "internes à l'arrêt", l'éclatement d'un arrêt multiple auquel passent n lignes nécessite (si l'on suppose $|L_r^-| = |L_r^+| = n$) au total $(2^n + 2n + 1)$ sommets ($(2^n - 1)$ sommets de répartition, n sommets d'embarquement, n sommets de descente, un sommet auxiliaire et un sommet d'attente) et $(n\,2^{n-1} + 2^n + 2n)$ arcs "internes" ($n\,2^{n-1}$ arcs de charge, (2^n-1) arcs d'attente, n arcs auxiliaires, n arcs de descente et un arc de correspondance). Il est donc évident que si le réseau étudié comporte des arrêts auxquels passent un très grand nombre de lignes, la taille du réseau éclaté risque d'être très grande, ce qui compliquerait la résolution du problème.

b) Les contraintes de répartition au sommet établissent des relations non-linéaires entre certains des flots, ce qui fait que le domaine des affectations réalisables n'est pas convexe (alors que dans le

cas où on a des équations de conservation de flot "normales", le domaine est toujours convexe). Il sera donc très difficile d'élaborer un algorithme nous permettant de trouver une affectation d'équilibre.

c) Pour une affectation réalisable donnée, le calcul du vecteur Π sera plus difficile à réaliser que dans le modèle pour les réseaux sans arrêts multiples : au lieu de n'avoir qu'à utiliser un algorithme de plus court chemin, il va falloir avoir recours à un algorithme de recherche de stratégies optimales conçu spécifiquement pour le problème.

d) Le sous-modèle d'arrêt multiple est beaucoup plus complexe à analyser que celui d'arrêt simple, comme nous le verrons dans le chapitre qui suit.

CHAPITRE SEPTIÈME

LE SOUS-MODÈLE D'ARRÊT MULTIPLE

7.1 Introduction

Le but de ce chapitre est l'élaboration d'un sous-modèle efficace pour déterminer la répartition des usagers entre les lignes attrayantes et pour calculer les temps d'attente aux arrêts multiples. Dans un premier temps, nous allons voir que l'on peut définir une classe de quasi-files d'attente extrêmement complexes qui permet de représenter adéquatement les arrêts multiples ; il nous sera cependant impossible de tirer des résultats utiles de l'étude de ces quasi-files. Par la suite, nous serons amenés à séparer en deux parties disjointes le problème qui nous intéresse : la première partie consistera à déterminer la répartition des usagers en fonction d'une certaine conjecture et la seconde à déduire de cette répartition et des paramètres de service des lignes les temps d'attente des usagers (en utilisant notamment les résultats du chapitre cinquième). Nous terminerons en montrant comment ces résultats peuvent être intégrés au modèle global d'équilibre du chapitre sixième.

7.2 Une classe de quasi-files d'attente pour les arrêts multiples

Dans le chapitre cinquième (section 5.2), nous avons vu que le phénomène d'attente des usagers à un arrêt simple pouvait être représenté par une quasi-file d'attente avec service en groupes, quasi-file d'attente

dans laquelle les arrivées de "clients" dans le système correspondent aux arrivées des usagers à l'arrêt et les intervalles de service (ou inter-service) aux intervalles interautobus sur la ligne desservant l'arrêt. Si l'on essaie d'appliquer le même type de modélisation aux arrêts multiples, on remarque tout d'abord que l'on a affaire dans ce cas à un système à plusieurs serveurs : à chaque ligne desservant l'arrêt (sauf celles dont c'est un terminus d'arrivée) correspondra évidemment un serveur ; qui plus est, les lignes ayant des paramètres de service différents, il est clair que nos serveurs sont hétérogènes. Pour ce qui est des "clients", on sait que les usagers ne retiennent qu'un sous-ensemble de lignes attrayantes ; en termes de quasi-files d'attente, cela revient à dire que les clients n'acceptent d'être servis que par certains serveurs ; on peut donc parler de clients sélectifs. On peut formaliser par une définition le concept que nous venons d'esquisser.

Définition 7.1 : Un système d'attente (file ou quasi-file d'attente) est dit à serveurs hétérogènes et à clients sélectifs si :

a) les serveurs ne sont pas identiques du point du vue des clients ;

b) les clients peuvent être partitionnés en un certain nombre de types correspondant à des sous-ensembles non-vides de l'ensemble des serveurs ;

c) pour chaque type de clients, ceux-ci n'acceptent d'être servis que par les serveurs appartenant au sous-ensemble associé à leur type.

On peut maintenant résumer la discussion précédente en disant qu'un arrêt multiple peut être modélisé par une quasi-file d'attente à serveurs hétérogènes et à clients sélectifs avec service en groupes. A notre connaissance, aucun article n'a jamais été publié sur un tel sujet ; c'est

donc dire qu'il nous faudra dériver nous-mêmes tous les résultats qui nous seront nécessaires.

Quand nous avons étudié les quasi-files d'attente au chapitre cinquième, nous avons passé sous silence la question de la discipline à l'intérieur de la file, car cela n'avait aucun impact sur le temps moyen d'attente (c'est un résultat bien connu) ; pour les quasi-files à serveurs hétérogènes et à clients sélectifs, la situation est cependant différente : comme les clients ont des attributs différents, on ne peut négliger le comportement dans la file. En effet, si un certain nombre d'usagers de différents types attendent à l'arrêt et qu'un autobus passe, la composition de la file après le passage de l'autobus dépendra non seulement de la composition que l'on avait avant le passage de l'autobus et de la ligne à laquelle appartient cet autobus, mais aussi du type des usagers qui auront réussi à monter dans l'autobus. De toute évidence, cela dépendra de la discipline de la file.

La discipline la plus commune dans les systèmes d'attente est la discipline "premier arrivé-premier servi" (PAPS ou FIFO). Si on analyse un arrêt multiple en supposant une telle discipline, on peut faire les constatations suivantes :

1) Pour décrire la composition de la file à un instant donné, il faut spécifier le type du client dans chaque position de la file.

2) Au moment du passage d'un autobus d'une ligne ℓ donnée avec K places libres, il est "facile" de déduire la composition de la file après le passage de l'autobus de celle existant avant ce

passage (il suffit de parcourir la file et de prendre dans celle-ci les K premiers clients pour lesquels la ligne ℓ est "attrayante").

3) Pour un arrêt avec M types de clients, il esixte M^N compositions possibles pour une file de N personnes (par exemple, pour un arrêt à trois lignes et sept types d'usagers, il existe 7^{20} compositions possibles pour une file de vingt personnes).

Pour faire une analyse en régime stationnaire en utilisant la technique des chaînes de Markov induites (comme l'on fait Bailey (1954) et Downton (1955, 1956) pour les quasi-files d'attente "simples"), il faut être en mesure de spécifier l'état du système étudié à certains instants choisis et pouvoir déterminer les transitions possibles (et les probabilités correspondantes) entre deux de ces instants successifs. Or, il est clair que, dans la description de l'état du système à un instant donné, il faut, entre autres choses, inclure une description complète de la composition de la file d'attente. De ceci, il ressort, à la lumière des remarques 1 et 3, qu'il est, à toutes fins pratiques, presqu'impossible de trouver une manière simple et efficace de définir l'ensemble des états pour une chaîne de Markov induite décrivant le système qui nous intéresse, et, conséquemment, qu'une analyse théorique directe des quasi-files d'attente à serveurs hétérogènes et à clients sélectifs avec service en groupes est une entreprise des plus ardues si l'on travaille avec la discipline PAPS.

Pour échapper aux difficultés relatives à la description de la file d'attente, on pourrait utiliser la discipline "anarchique", c'est-à-dire

une discipline dans laquelle l'ordre d'arrivée n'a aucune importance. Dans ce cas, on remarque que :

1) Pour décrire la composition de la file à un instant donné, il suffit de spécifier le nombre total d'usagers de chaque type ; c'est donc dire que la composition de la file est donnée par un vecteur de dimension M où M est le nombre de types de clients.

2) Il est assez laborieux de déterminer les usagers qui vont prendre place dans un autobus qui passe à l'arrêt, car la file n'a plus de structure ordonnée. Plus précisément, le processus de modification de la file d'attente, au lieu d'être déterministe, comme dans le cas de la discipline PAPS, est maintenant aléatoire : il faut choisir au hasard les usagers qui vont occuper les places libres, parmi ceux qui désirent monter dans l'autobus en question.

En changeant de discipline pour la file d'attente, on a donc énormément simplifié le problème de la description de l'état du système à un instant donné, mais, en contrepartie, il est maintenant beaucoup plus difficile de déterminer les probabilités de transition entre les états (notons ici que, pour construire la matrice de transition de la chaîne de Markov induite associée au processus qui nous intéresse, il ne suffit pas de tenir compte des transitions qui s'effectuent au moment des passages d'autobus, mais aussi des arrivées d'usagers entre les passages et d'autres éléments dépendant des distributions choisies pour les intervalles interautobus). Evidemment, il serait possible, avec beaucoup de travail, de pousser plus loin cette analyse pour les cas les plus simples (arrêts desservis par seulement deux lignes, temps de service exponentiels),

mais, devant les difficultés soulevées par une telle approche pour les cas plus compliqués, il est sans doute judicieux de ne pas s'y attarder davantage et d'essayer, plutôt, d'attaquer le problème sous un angle différent.

7.3 Séparation du problème de répartition et du problème d'attente

Nous avons vu dans la dernière section qu'une approche "directe" du problème de répartition et d'attente des usagers aux arrêts multiples par les modèles de quasi-files d'attente semblait impossible. Pour contourner cette difficulté, on peut "séparer" notre problème en deux parties : dans un premier temps, on essaiera de déterminer par une méthode approximative la répartition des usagers entre les lignes attrayantes ; par la suite, connaissant cette répartition, on définira pour chaque type d'usagers une quasi-file d'attente dont le temps d'attente moyen en régime stationnaire constituera une approximation du temps d'attente moyen des usagers du type considéré ; pour définir les fonctions d'attente associées à chaque type d'usagers, il ne nous restera plus qu'à utiliser les résultats obtenus dans le chapitre cinquième.

Il est certain qu'en séparant le problème original, on perdra une partie de l'information disponible, mais il y aura aussi économie d'efforts et le modèle final obtenu sera plus facilement "manipulable" puisque les fonctions d'attente pour les arrêts simples et pour les arrêts multiples auront des formes similaires.

7.4 Sous-modèle de répartition

7.4.1 Introduction

Dans cette section, nous allons définir un modèle nous permettant de déterminer la répartition des usagers à un arrêt multiple r donné, considéré isolément. Ce modèle est basé sur une hypothèse fondamentale appelée "hypothèse de répartition au prorata de la capacité résiduelle" qui sera explicitée à la sous-section 7.4.4.

7.4.2 Notation

La notation que nous utiliserons dans cette section sera, en majeure partie, la même que dans les chapitres précédents. Nous laisserons cependant tomber tous les indices "r" relatifs à l'arrêt considéré, puisque nous n'étudions qu'un arrêt à la fois.

Nous poserons donc :

L^+, ensemble des lignes pour lesquelles l'arrêt n'est pas un terminus d'arrivée ;

μ_ℓ, fréquence de la ligne ℓ, $\ell \in L^+$;

Δ, ensemble des types d'usagers ;

λ_δ, taux d'arrivée des usagers du type δ, $\delta \in \Delta$

$\overline{K}_\ell$, nombre moyen de places libres (à l'arrêt r) pour les autobus de la ligne ℓ, $\ell \in L^+$;

$\nu_\ell \equiv \mu_\ell \overline{K}_\ell$, taux de service de la ligne ℓ, $\ell \in L^+$;

$\Delta^+ \equiv \{\delta \in \Delta \mid \lambda_\delta > 0\}$, ensemble des types "actifs" d'usagers ;

$\Delta^+(\ell) \equiv \{\delta \epsilon \Delta^+ | \ell \epsilon \delta\}$, ensemble des types d'usagers utilisant la ligne δ.

Rappelons ici que l'ensemble des types d'usagers correspond à l'ensemble des décisions possibles à l'arrêt r et donc que $\Delta = P(L^+)-\{\emptyset\}$.

7.4.3 Condition de stationnarité

Dans le chapitre cinquième, nous avons vu qu'une quasi-file d'attente $Q\text{-}M/E_k^{[K]}/1$ possédait un état stationnaire si son coefficient d'achalandage était strictement inférieur à 1 ; dans cette sous-section, nous allons énoncer une condition équivalente pour le modèle d'arrêt multiple (cette condition pouvant être généralisée à tous les modèles d'attente à serveurs hétérogènes et à clients sélectifs).

Dans un modèle à serveur unique, on n'a pas d'état stationnaire si le mécanisme de service est saturé, c'est-à-dire si le taux d'arrivée des usagers est plus grand ou égal au taux de service. Si on applique le même raisonnement à chacun des serveurs d'un système à serveurs hétérogènes, il est clair que, si le taux d'arrivée des clients qui exigent d'être servis par un serveur donné excède la capacité de service de celui-ci, on ne pourra atteindre un état stationnaire. Le système ne pourra non plus posséder d'état stationnaire s'il existe un sous-ensemble de serveurs dont le taux de service combiné est inférieur au taux d'arrivée combiné des usagers qui ne peuvent être servis que par ces serveurs. Ce raisonnement nous amène au théorème suivant.

Théorème 7.1

Un système d'attente à serveurs hétérogènes et à clients sélectifs dans lequel les distribution des intervalles interarrivées et des intervalles de service sont des lois d'Erlang possède un état stationnaire si et seulement si le taux de service combiné de tout sous-ensemble de serveurs est strictement supérieur à la somme des taux d'arrivée des types d'usagers inclus dans ce sous-ensemble.

Démonstration

Montrons tout d'abord que la condition est nécessaire. Il est évident que, si un sous-ensemble de serveurs est saturé, le système ne pourra tendre vers un état stationnaire, car la file d'attente des usagers des types inclus dans ce sous-ensemble va devenir infinie.

Il est plus difficile de montrer que la condition énoncée est une condition suffisante. Sans entrer dans tous les détails, on peut remarquer que, si la condition proposée est vérifiée pour tout sous-ensemble de serveurs, il est possible de répartir les usagers entre les serveurs de sorte que le coefficient d'achalandage de chaque serveur soit inférieur à 1. Ceci nous assure donc que la file d'attente ne tendra pas vers l'infini. Le système, bien que n'exhibant pas un comportement "explosif", pourrait ne pas posséder d'état stationnaire si le phénomène étudié était de nature périodique. Heureusement, tel ne sera pas le cas ici, étant donné que les distributions des intervalles interarrivées et des intervalles de service sont des Erlang.

□

Dans l'énoncé du théorème, nous nous sommes restreints à ne considérer que les systèmes d'attente décrits par des lois d'Erlang. A la lecture de la démonstration, il est clair que le théorème pourrait être généralisé à tous les systèmes d'attente qui n'ont pas un comportement périodique.

Avec la notation définie à la sous-section précédente, la condition de stationnarité peut être réécrite comme suit :

$$\sum_{\delta' \subset \delta} \lambda_{\delta'} < \sum_{\ell \in \delta} \nu_\ell , \qquad \forall\ \delta \in \Delta.$$

Dans la suite, nous ferons toujours l'hypothèse que le système étudié vérifie la condition de stationnarité.

7.4.4 Hypothèse fondamentale de répartition

On voudrait trouver un principe qui nous permette de répartir les usagers entre les lignes d'une manière qui corresponde assez bien à la réalité et qui, d'autre part, soit relativement simple.

Pour un système assez congestionné, il semble raisonnable de penser que c'est la disponibilité des places libres sur chaque ligne qui soit le facteur déterminant dans la répartition. D'autre part, il y a compétition entre les divers types d'usagers pour l'accès aux autobus, ce qui nous amène à formuler notre principe de répartition en fonction de la capacité résiduelle des lignes plutôt que de la capacité brute de celles-ci. Nous proposons donc :

Hypothèse de répartition au prorata de la capacité résiduelle :

Pour chaque type d'usagers, ceux-ci se répartissent entre les lignes attrayantes pour eux au prorata de la capacité résiduelle de celles-ci.

Par "capacité résiduelle" d'une ligne, on peut entendre soit la capacité restant pour la ligne, une fois que les usagers des autres types ont été répartis entre les lignes, soit tout simplement la différence entre la capacité totale de la ligne (à cet arrêt) et la charge d'usagers qui lui est affectée à l'arrêt. Nous montrerons dans la sous-section qui suit que les deux interprétations sont équivalentes, dans le sens qu'elles peuvent s'exprimer mathématiquement par des systèmes d'équations équivalentes.

7.4.5 Répartitions réalisables et répartitions d'équilibre

Dans cette sous-section, nous allons caractériser les répartitions des usagers qui sont possibles ("réalisables") et celles qui satisfont l'hypothèse de répartition au prorata de la capacité résiduelle (nous dirons que ce sont des répartitions d'équilibre).

Définition 7.2

Soit ω_ℓ^δ le taux d'usagers de type δ prenant la ligne ℓ à un arrêt donné, alors la répartition $\Omega(\equiv(\omega_\ell^\delta)_{\ell\in L^+}^{\delta\in\Delta^+})$ est dite réalisable si elle satisfait le système d'équations et d'inéquations suivant :

$$\sum_{\ell \in \delta} \omega_\ell^\delta = \lambda_\delta \quad , \qquad \forall \ \delta \in \Delta^+ \ ,$$

$$\sum_{\delta \in \Delta^+(\ell)} \omega_\ell^\delta < \nu_\ell \quad , \qquad \forall \ \ell \in L^+ \ .$$

$$\omega_\ell^\delta \geq 0 \quad , \qquad \forall \ \delta \in \Delta^+ \ , \ \forall \ \ell \in L^+$$

Une répartition est donc réalisable si tous les usagers sont affectés à une ligne attrayante et si, en conséquence de cette affectation, aucune ligne n'est saturée.

Définition 7.3

Une répartition réalisable Ω est dite répartition d'équilibre si elle satisfait l'hypothèse de répartition au prorata de la capacité résiduelle, à savoir

$$\omega_\ell^\delta = \lambda_\delta \left(\frac{\nu_\ell - \sum_{\substack{\delta' \in \Delta^+(\ell) \\ \delta' \neq \delta}} \omega_\ell^{\delta'}}{\sum_{\ell' \in \delta} \left(\nu_{\ell'} - \sum_{\substack{\delta' \in \Delta^+(\ell') \\ \delta' \neq \delta}} \omega_{\ell'}^{\delta'} \right)} \right) \ , \ \forall \ \delta \in \Delta^+, \ \forall \ \ell \in L^+ \quad (1)$$

ou

$$\omega_\ell^\delta = \lambda_\delta \left(\frac{\nu_\ell - \sum_{\delta' \in \Delta^+(\ell)} \omega_\ell^{\delta'}}{\sum_{\ell' \in \delta} \left(\nu_{\ell'} - \sum_{\delta' \in \Delta^+(\ell')} \omega_{\ell'}^{\delta'} \right)} \right) \ , \ \forall \ \delta \in \Delta^+, \ \forall \ \ell \in L^+ \quad (2)$$

La première équation correspond à la première interprétation de la notion de "capacité résiduelle" et la seconde à la seconde interprétation. Nous allons maintenant montrer qu'elles sont équivalentes. Pour ce faire, appelons w_ℓ la "capacité libre" de la ligne ℓ, capacité définie comme suit :

$$w_\ell = v_\ell - \sum_{\delta' \in \Lambda^+(\ell)} \lambda_{\delta'}, \qquad \forall\ \ell \in L^+. \tag{3}$$

On peut alors réécrire les équations (1) et (2) sous les formes équivalentes (1') et (2') :

$$\omega_\ell^\delta = \lambda_\delta \left(\frac{w_\ell + \omega_\ell^\delta}{\sum_{\ell' \in \delta} (w_{\ell'} + \omega_{\ell'}^\delta)} \right), \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+ \tag{1'}$$

$$\omega_\ell^\delta = \lambda_\delta \left(\frac{w_\ell}{\sum_{\ell' \in \delta} w_{\ell'}} \right), \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+. \tag{2'}$$

Maintenant, on peut remarquer que

$$(1') \iff \omega_\ell^\delta = \lambda_\delta \left(\frac{w_\ell + \omega_\ell^\delta}{\sum_{\ell' \in \delta} w_{\ell'} + \sum_{\ell' \in \delta} \omega_{\ell'}^\delta} \right), \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+$$

$$\iff \omega_\ell^\delta = \lambda_\delta \left(\frac{w_\ell + \omega_\ell^\delta}{\lambda_\delta + \sum_{\ell' \in \delta} w_{\ell'}} \right), \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+$$

$$\left(\text{par } \lambda_\delta = \sum_{\ell' \in \delta} \omega_{\ell'}^\delta \right)$$

$$\iff \omega_\ell^\delta (\lambda_\delta + \sum_{\ell' \in \delta} w_{\ell'}) = \lambda_\delta (w_\ell + \omega_\ell^\delta), \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+$$

$$\iff \omega_\ell^\delta \sum_{\ell' \in \delta} w_{\ell'} = \lambda_\delta w_\ell, \quad \forall\ \delta \in \Delta^+, \ \forall\ \ell \in L^+$$

$$(1') \iff (2').$$

Ceci démontre bien l'équivalence des deux interprétations.

Nous allons maintenant écrire un système d'équations et d'inéquations qui résume les conditions que doit satisfaire une répartition d'équilibre. Ce système est obtenu en regroupant les conditions définissant les répartitions réalisables et les équations (2') et (3).

$$(I)\quad \begin{cases} \sum\limits_{\ell\in\delta} \omega_\ell^\delta = \lambda_\delta & , \ \forall\ \delta\in\Delta^+ \\ w_\ell > 0 & , \ \forall\ \ell\in L^+ \\ \omega_\ell^\delta \geq 0 & , \ \forall\ \delta\in\Delta^+,\ \forall\ \ell\in L^+ \\ w_\ell = \nu_\ell - \sum\limits_{\delta'\in\Delta^+(\ell)} \omega_\ell^{\delta'} & , \ \forall\ \ell\in L^+ \\ \omega_\ell^\delta = \lambda_\delta \left(\dfrac{w_\ell}{\sum\limits_{\ell'\in\delta} w_{\ell'}} \right) & , \ \forall\ \delta\in\Delta^+ ,\ \forall\ \ell\in L^+ \end{cases}$$

On peut remarquer immédiatement que, si les autres équations et inéquations sont satisfaites, les équations $\sum\limits_{\ell\in\delta} \omega_\ell^\delta = \lambda_\delta$ et les inéquations $\omega_\ell^\delta \geq 0$ sont redondantes. On peut donc remplacer le système (I) par le système équivalent

$$(II)\quad \begin{cases} w_\ell = \nu_\ell - \sum\limits_{\delta'\in\Delta^+} \omega_\ell^{\delta'} & , \ \forall\ \ell\in L^+ \\ w_\ell > 0 & , \ \forall\ \ell\in L^+ \\ \omega_\ell^\delta = \lambda_\delta \left(\dfrac{w_\ell}{\sum\limits_{\ell'\in\delta} w_{\ell'}} \right) & , \ \forall\ \delta\in\Delta^+,\ \forall\ \ell\in L^+ \end{cases}$$

A l'origine, les w_ℓ ne jouaient qu'un rôle de variables auxiliaires dans ce sytème, mais on peut ramener notre problème à la recherche de ces seules variables. En effet, si on connaît les valeurs des w_ℓ, chaque ω_ℓ^δ est déterminé de manière unique. Si, dans la première série d'équations, on remplace les $\omega_\ell^{\delta'}$ par leur expression en fonction des w_ℓ, on obtient le système

$$(III)\begin{cases} w_\ell = \nu_\ell - \sum\limits_{\delta' \in \Delta^+(\ell)} \lambda_{\delta'} \left(\dfrac{w_\ell}{\sum\limits_{\ell' \in \delta'} w_{\ell'}} \right) , & \forall\ \ell \in L^+ \quad (4) \\ w_\ell \geq 0 , & \forall\ \ell \in L^+ \quad (5) \\ \omega_\ell^\delta = \lambda_\delta \left(\dfrac{w_\ell}{\sum\limits_{\ell' \in \delta} w_{\ell'}} \right) , & \forall\ \delta \in \Delta^+,\ \forall\ \ell \in L^+ \quad (6) \end{cases}$$

Pour résoudre le système (III), il suffit de résoudre le sous-système (IV) formé des équations (4) et des inéquations (5). Il est intéressant de noter que le système (IV) ne comporte que $|L^+|$ équations et $|L^+|$ inéquations en $|L^+|$ variables, alors que si on avait décidé se travailler avec les ω_ℓ^δ, il nous aurait fallu au minimum $|L^+| \times |\Delta^+|$ équations (une par ω_ℓ^δ) et $|L^+|$ inéquations (une par ligne) en $|L^+| \times |\Delta^+|$ variables.

Nous allons maintenant modifier légèrement le système (IV) pour mettre en évidence certaines de ses propriétés. On a

$$(4) \quad \Leftrightarrow w_\ell + \sum_{\delta \in \Delta^+(\ell)} \lambda_\delta \left[\frac{w_\ell}{\sum_{\ell' \in \delta} w_{\ell'}} \right] = v_\ell \ , \ \forall \ \ell \in L^+$$

$$\Leftrightarrow w_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}} \right] = v_\ell \ , \quad \forall \ \ell \in L^+$$

$$\Leftrightarrow w_\ell = v_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}} \right]^{-1} , \quad \forall \ \ell \in L^+ \qquad (7)$$

Le système (IV) est donc équivalent au système (V) suivant :

$$(V) \quad \begin{cases} w_\ell = v_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}} \right]^{-1} , & \forall \ \ell \in L^+ \\ w_\ell > 0 , & \forall \ \ell \in L^+ \end{cases}$$

Pour trouver une répartition d'équilibre Ω pour un arrêt multiple donné, il suffit donc de résoudre le système (V), puis de déterminer les ω_ℓ^2 en utilisant l'équation (6).

7.4.6 Existence et unicité de la répartition d'équilibre

Dans cette sous-section, nous allons montrer que, pour un arrêt multiple satisfaisant la condition de stationnarité, il existe toujours une répartition d'équilibre unique.

Théorème 7.2 (Théorème d'existence)

Pour un arrêt multiple satisfaisant la condition de stationnarité, il existe au moins une répartition d'équilibre.

Démonstration

Pour démontrer l'existence d'une répartition d'équilibre, il suffit de montrer que le système

$$\begin{cases} w_\ell = \nu_\ell \left[1 + \sum\limits_{\delta \in \Delta^+(\ell)} \dfrac{\lambda_\delta}{\sum\limits_{\ell' \in \delta} w_{\ell'}} \right]^{-1} , & \forall\ \ell \in L^+ \\ w_\ell > 0 , & \forall\ \ell \in L^+ \end{cases}$$

possède au moins une solution.

Remarquons tout d'abord que l'hypothèse de stationnarité implique que $\nu_\ell > 0$, $\forall\ \ell \in L^+$.

Notons aussi que, si jamais on avait $\Delta^+(\ell) = \emptyset$ pour un ℓ quelconque, il suffirait de poser $w_\ell = \nu_\ell$ pour ce ℓ et de l'ignorer par la suite. Pour alléger la preuve, nous supposerons dans la suite que $\Delta^+(\ell) \neq \emptyset$, $\forall\ \ell \in L^+$.

Posons maintenant $\underline{w}^{(0)} = \underline{\nu}$ et construisons la séquence $\{\underline{w}^{(i)}\}$, i=1,2,..., par la règle

$$w_\ell^{(i+1)} = \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum\limits_{\ell' \in \delta} w_{\ell'}} \right]^{-1} , \qquad \forall\ \ell \in L^+$$

Montrons tout d'abord que $0 < \underline{w}^{(i+1)} < \underline{w}^{(i)}$, i=0,1,2,... Pour ce faire, on peut procéder par récurrence. On a pour i=0,

$$w_\ell^{(1)} = \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum\limits_{\ell' \in \delta} \nu_{\ell'}} \right]^{-1} , \qquad \forall\ \ell \in L^+$$

$$\Rightarrow \quad 0 < w_\ell^{(1)} < \nu_\ell = w_\ell^{(0)} , \qquad \forall\ \ell \in L^+,$$

et donc la propriété est vérifiée.

Supposons maintenant que la propriété est vraie pour i =0,1,...,j et montrons qu'elle l'est aussi pour i =j +1.

On a

$$w_{\ell}^{(j+1)} = v_{\ell}\left[1 + \sum_{\delta \in \Delta^{+}(\ell)} \frac{\lambda_{\delta}}{\sum_{\ell' \in \delta} w_{\ell'}^{(j)}}\right]^{-1}, \qquad \forall\ \ell \in L^{+}$$

$$\Rightarrow \quad w_{\ell}^{(j+1)} < v_{\ell}\left[1 + \sum_{\delta \in \Delta^{+}(\ell)} \frac{\lambda_{\delta}}{\sum_{\ell' \in \delta} w_{\ell'}^{(j-1)}}\right]^{-1} \qquad \forall\ \ell \in L^{+}$$

puisque $\underline{w}^{(j)} < \underline{w}^{(j-1)}$ par l'hypothèse de récurrence.

Et donc, $w_{\ell}^{(j+1)} < w_{\ell}^{(j)}$, $\forall\ \ell \in L^{+}$.

D'autre part, si $\underline{w}^{(j)} > 0$, il est évident que $\underline{w}^{(j+1)} > 0$. La propriété $0 < \underline{w}^{(i+1)} < \underline{w}^{(i)}$, i=0,1,2,..., est donc démontrée.

Remarquons que ce résultat implique que

$$w_{\ell}^{(i)} > v_{\ell}\left[1 + \sum_{\delta \in \Delta^{+}(\ell)} \frac{\lambda_{\delta}}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}}\right]^{-1}, \qquad \forall\ \ell \in L^{+}$$

$$\Rightarrow \quad w_{\ell}^{(i)}\left[1 + \sum_{\delta \in \Delta^{+}(\ell)} \frac{\lambda_{\delta}}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}}\right] > v_{\ell}, \qquad \forall\ \ell \in L^{+}$$

$$\Rightarrow \quad \sum_{\ell \in L^{+}} \left\{ w_{\ell}^{(i)}\left[1 + \sum_{\delta \in \Delta^{+}(\ell)} \frac{\lambda_{\delta}}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}}\right]\right\} > \sum_{\ell \in L^{+}} v_{\ell}$$

$$\Rightarrow \quad \sum_{\ell \in L^+} w_\ell^{(i)} + \sum_{\ell \in L^+} \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta w_\ell^{(i)}}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}} > \sum_{\ell \in L^+} \nu_\ell$$

$$\Rightarrow \quad \sum_{\ell \in L^+} w_\ell^{(i)} + \sum_{\delta \in \Delta^+} \lambda_\delta \frac{\sum_{\ell \in \delta} w_\ell^{(i)}}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}} > \sum_{\ell \in L^+} \nu_\ell$$

$$\Rightarrow \quad \sum_{\ell \in L^+} w_\ell^{(i)} > \sum_{\ell \in L^+} \nu_\ell - \sum_{\delta \in \Delta^+} \lambda_\delta$$

et ce pour tout i.

Soit $W = \{\underline{w} \in R^{|L^+|} \mid 0 < \underline{w} \leq \underline{\nu} \text{ et } \sum_{\ell \in L^+} w_\ell > \sum_{\ell \in L^+} \nu_\ell - \sum_{\delta \in \Delta^+} \lambda_\delta\}$

et soit $\overline{W}$ sa fermeture. Il est évident que

$$\overline{W} = \{\underline{w} \in R^{|L^+|} \mid 0 \leq \underline{w} \leq \underline{\nu} \text{ et } \sum_{\ell \in L^+} w_\ell \geq \sum_{\ell \in L^+} w_\ell - \sum_{\delta \in \Delta^+} \lambda_\delta\}$$

On a $\underline{w}^{(i)} \in W$, i=0,1,2,..., et la suite $\{\underline{w}^{(i)}\}_{i=0}^{\infty}$ possède un point d'accumulation $\underline{w}^*$ dans $\overline{W}$. Il convient de remarquer ici que $\underline{w}^*$ est la limite de la suite $\{\underline{w}^{(i)}\}_{i=0}^{\infty}$ elle-même, car, pour chaque indice ℓ, la suite $\{w_\ell^{(i)}\}_{i=0}^{\infty}$ est une suite monotone strictement décroissante, bornée vers le bas, et par conséquent convergente.

Observons maintenant ce qui se passe au point $\underline{w}^*$.

Soit $\overline{L}^+ = \{\ell \in L^+ \mid w_\ell^* = 0\}$

et $\tilde{L}^+ = \{\ell \in L^+ \mid w_\ell^* > 0\}$.

Pour $\ell \in \tilde{L}^+$, on a

$$w_\ell^* = \lim_{i \to \infty} w_\ell^{(i+1)}$$

$$= \lim_{i \to \infty} \left\{ \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}^{(i)}} \right]^{-1} \right\}$$

$$= \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\lim_{i \to \infty} \sum_{\ell' \in \delta} w_{\ell'}^{(i)}} \right]^{-1}$$

$$\therefore \quad w_\ell^* = \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}^*} \right]^{-1}$$

(Remarque : la limite ne pose aucun problème ici, car

$\sum_{\ell' \in \delta} w_{\ell'}^* \geq w_\ell^* > 0, \quad \forall \ \delta \in \Delta^+(\ell)$)

De ceci, on tire

$$w_\ell^* \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}^*} \right] = \nu_\ell, \quad \forall \ \ell \in \tilde{L}^+$$

ce qui peut être réécrit sous la forme

$$w_\ell^* \left[1 + \frac{\lambda_\delta}{\sum_{\ell' \in \delta \cap \tilde{L}^+} w_{\ell'}^*} \right] = \nu_\ell \quad , \quad \forall \ \ell \in \tilde{L}^+,$$

puisque $w_{\ell'}^* = 0, \quad \forall \ \ell' \notin \tilde{L}^+$.

Faisant la somme sur tous les éléments de $\tilde{L}^+$, on obtient

$$\sum_{\ell \in \tilde{L}^+} w_\ell^* \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta \cap \tilde{L}^+} w_{\ell'}^*} \right] = \sum_{\ell \in \tilde{L}^+} v_\ell$$

$$\Rightarrow \quad \sum_{\ell \in \tilde{L}^+} w_\ell^* + \sum_{\ell \in \tilde{L}^+} \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta w_\ell^*}{\sum_{\ell' \in \delta \cap \tilde{L}^+} w_{\ell'}^*} = \sum_{\ell \in \tilde{L}^+} v_\ell$$

Posons $\tilde{\Delta}^+ = \bigcup_{\ell \in \tilde{L}^+} \Delta^+(\ell)$. On obtient alors

$$\sum_{\ell \in \tilde{L}^+} w_\ell^* + \sum_{\delta \in \tilde{\Delta}^+} \lambda_\delta \frac{\sum_{\ell \in \delta \cap \tilde{L}^+} w_\ell^*}{\sum_{\ell' \in \delta \cap \tilde{L}^+} w_{\ell'}^*} = \sum_{\ell \in \tilde{L}^+} v_\ell$$

$$\Rightarrow \quad \sum_{\ell \in \tilde{L}^+} w_\ell^* = \sum_{\ell \in \tilde{L}^+} v_\ell - \sum_{\delta \in \tilde{\Delta}^+} \lambda_\delta \qquad (8)$$

D'autre part,

$$\underline{w}^* \in \overline{W} \Rightarrow \sum_{\ell \in L^+} w_\ell^* \geq \sum_{\ell \in L^+} v_\ell - \sum_{\delta \in \Delta^+} \lambda_\delta \qquad (9)$$

Soustrayant membre à membre l'équation (8) de l'inéquation (9), on obtient

$$\sum_{\ell \in L^+} w_\ell^* - \sum_{\ell \in \tilde{L}^+} w_\ell^* \geq \left(\sum_{\ell \in L^+} v_\ell - \sum_{\delta \in \Delta^+} \lambda_\delta \right) - \left(\sum_{\ell \in \tilde{L}^+} v_\ell - \sum_{\delta \in \tilde{\Delta}^+} \lambda_\delta \right)$$

$$\Rightarrow \sum_{\ell \in L^+} w_\ell^* \geq \sum_{\ell \in L^+} v_\ell - \sum_{\delta \in \Delta^+ - \tilde{\Delta}^+} \lambda_\delta$$

puisque $L^+ - \tilde{L}^+ = \overline{L}^+$.

$$\Rightarrow \sum_{\delta \in \Delta^+ - \tilde{\Delta}^+} \lambda_\delta \geq \sum_{\ell \in \bar{L}^+} \nu_\ell$$

car $w^*_\ell = 0$, $\forall\ \ell \in \bar{L}^+$.

Mais $\Delta^+ - \tilde{\Delta}^+ = \Delta^+ - \left[\bigcup_{\ell \in \tilde{L}^+} \Delta^+(\ell) \right]$

$$= \{\delta \in \Delta^+ \mid \delta \subseteq L^+ - \tilde{L}^+\}$$

$$\Delta^+ - \tilde{\Delta}^+ = \{\delta \in \Delta^+ \mid \delta \subseteq \bar{L}^+\}$$

On a donc

$$\sum_{\delta \subseteq \bar{L}^+} \lambda_\delta \geq \sum_{\ell \in \bar{L}^+} \nu_\ell$$

Mais, par l'hypothèse de stationnarité, on a

$$\sum_{\delta'' \subseteq \delta'} \lambda_{\delta''} < \sum_{\ell \in \delta'} \nu_\ell , \quad \forall\ \delta' \in \Delta ,$$

c'est-à-dire pour tout sous-ensemble non-vide δ' de L^+. On en conclut donc que $\bar{L}^+ = \emptyset$, c'est-à-dire que $\tilde{L}^+ = L^+$.

En définitive, on peut affirmer que

$$w^*_\ell > 0,$$

et que $$w^*_\ell = \nu_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w^*_{\ell'}} \right]^{-1} , \quad \forall\ \ell \in L^+,$$

ce qui démontre que le système d'équations et d'inéquations posé au départ possède au moins une solution ($\underline{w}^*$) et, donc, qu'il existe au moins une répartition d'équilibre pour l'arrêt.

□

Théorème 7.3 (Théorème d'unicité)

Pour un arrêt multiple satisfaisant la condition de stationnarité, il n'existe qu'une seule répartition d'équilibre

Démonstration

Pour démontrer l'unicité de la répartition d'équilibre, il suffit de montrer que le système

$$\begin{cases} w_\ell = v_\ell \left[1 + \sum_{\delta \in \Delta^+(\ell)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} w_{\ell'}} \right]^{-1} , & \forall \ \ell \in L^+ \\ w_\ell > 0 , & \forall \ \ell \in L^+ \end{cases}$$

ne possède qu'une solution.

Procédons par l'absurde et supposons que $\bar{w}$ et $\tilde{w}$ sont deux solutions distinctes de ce système.

Supposons aussi, sans perte de généralité, que $\bar{w}_\ell > \tilde{w}_\ell$ pour au moins un indice ℓ, et soit ℓ^* un indice tel que

$$\frac{\bar{w}_{\ell^*}}{\tilde{w}_{\ell^*}} = \max_{\ell \in L^+} \left(\frac{\bar{w}_\ell}{\tilde{w}_\ell} \right) .$$

(Remarque : tous les rapports $\frac{\bar{w}_\ell}{\tilde{w}_\ell}$ sont finis, car $\tilde{w}_\ell > 0$, $\forall \ \ell \in L^+$, par hypothèse.)

On a alors $\bar{w}_{\ell^*} > \tilde{w}_{\ell^*}$.

D'autre part, puisque $\bar{w}$ et $\tilde{w}$ sont des solutions du système d'équations et d'inéquations, on a

$$\bar{w}_{\ell^*} \left[1 + \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \bar{w}_{\ell'}} \right] = v_{\ell^*}$$

et

$$\tilde{w}_{\ell^*} \left[1 + \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \tilde{w}_{\ell'}} \right] = v_{\ell^*} \ .$$

$$\Rightarrow (\bar{w}_{\ell^*} - \tilde{w}_{\ell^*}) + \bar{w}_{\ell^*} \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \bar{w}_{\ell'}} - \tilde{w}_{\ell^*} \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \tilde{w}_{\ell'}} = 0 \ .$$

Comme $\bar{w}_{\ell^*} - \tilde{w}_{\ell^*} > 0$, on en déduit que

$$\bar{w}_{\ell^*} \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \bar{w}_{\ell'}} < \tilde{w}_{\ell^*} \sum_{\delta \in \Delta^+(\ell^*)} \frac{\lambda_\delta}{\sum_{\ell' \in \delta} \tilde{w}_{\ell}}$$

Il existe donc $\delta^* \in \Delta^+(\ell^*)$ tel que

$$\bar{w}_{\ell^*} \frac{\lambda_{\delta^*}}{\sum_{\ell' \in \delta^*} \bar{w}_{\ell'}} < \tilde{w}_{\ell^*} \frac{\lambda_{\delta^*}}{\sum_{\ell' \in \delta^*} \tilde{w}_{\ell'}}$$

$$\Rightarrow \quad \frac{\bar{w}_{\ell^*}}{\tilde{w}_{\ell^*}} < \frac{\sum_{\ell' \in \delta^*} \bar{w}_{\ell'}}{\sum_{\ell' \in \delta^*} \tilde{w}_{\ell'}}$$

De ceci, on conclut qu'il doit exister $\ell^{**} \in \delta^*$ tel que

$$\frac{\bar{w}_{\ell**}}{\tilde{w}_{\ell**}} > \frac{\bar{w}_{\ell*}}{\tilde{w}_{\ell*}} ,$$

mais ceci est impossible puisque $\frac{\bar{w}_{\ell}}{\tilde{w}_{\ell}} \leq \frac{\bar{w}_{\ell*}}{\tilde{w}_{\ell*}}$, $\forall \ \ell \in L^{+}$, par construction.

On peut donc conclure que notre système d'équations et d'inéquations ne peut posséder qu'une seule solution et, par conséquent, qu'il n'existe qu'une seule répartition d'équilibre pour un arrêt multiple satisfaisant la condition de stationnarité.

□

Les deux théorèmes que nous venons de démontrer sont d'une extrême importance, car ils nous indiquent que l'hypothèse de répartition au prorata de la capacité résiduelle est "fructueuse" (pour les cas qui nous intéressent, on aura toujours une répartition d'équilibre unique). De plus, la preuve constructive du théorème 7.2 nous fournit implicitement un algorithme de calcul des répartitions d'équilibre. Il nous a cependant été impossible jusqu'à maintenant d'obtenir des résultats sur le taux de convergence de cet algorithme.

7.4.7 Détermination des probabilités de répartition ; propriétés

Une fois que l'on connaît une répartition d'équilibre Ω^* pour un arrêt donné, il est facile d'en tirer des probabilités de répartition (rappelons que ce sont des probabilités dont nous avons besoin pour le modèle global d'équilibre). En effet, il suffit de poser

$$p_{\ell}^{\delta} = \frac{\omega^{*\delta}_{\ell}}{\lambda_{\delta}} \quad , \qquad \ell \in \delta \ , \delta \in \Delta^{+} \ .$$

En fait, il est plus simple de définir directement les probabilités de répartition à partir de la solution $\underline{w^*}$ du système (V). On peut alors écrire

$$p_{\ell}^{\delta} = \frac{w^*_{\ell}}{\sum_{\ell' \in \delta} w^*_{\ell'}} \quad , \qquad \ell \in \delta \ , \delta \in \Delta \ .$$

Notons qu'avec cette définition directe, il devient possible d'écrire des probabilités de répartition même pour les types d'usagers avec taux d'arrivée nul, ce qui pourrait être utile dans la recherche d'une affectation d'équilibre pour le modèle global.

Il faut souligner une propriété importante des probabilités de répartition obtenues à partir de l'hypothèse de répartition au prorata de la capacité résiduelle, propriété que nous appellerons "propriété d'indépendance des probabilités conditionnelles". Pour un usager de type δ', on appellera "probabilité conditionnelle de prendre la ligne ℓ étant donné qu'une ligne du sous-ensemble δ est choisie" (avec $\ell \in \delta$ et $\delta \subset \delta'$) la probabilité $\bar{p}^{\delta'}_{\ell|\delta}$ définie comme suit :

$$\bar{p}^{\delta'}_{\ell|\delta} = \frac{p_{\ell}^{\delta'}}{\sum_{\ell' \in \delta} p_{\ell'}^{\delta'}} \quad , \qquad \forall \ \ell \in \delta \ , \forall \ \delta \subset \delta' \qquad (10)$$

Si l'on remplace dans (10), les $p_\ell^{\delta'}$ par leur valeur exprimée par rapport à $\underline{w}^*$, on obtient

$$\bar{p}_{\ell|\delta}^{\delta'} = \frac{\dfrac{w_\ell^*}{\sum_{\ell''\in\delta'} w_{\ell''}^*}}{\sum_{\ell'\in\delta} \dfrac{w_{\ell'}^*}{\sum_{\ell''\in\delta'} w_{\ell''}^*}}, \qquad \forall\ \ell\in\delta\ ,\ \forall\ \delta\subseteq\delta',\ \forall\ \delta'\in\Delta$$

$$\Rightarrow \qquad \bar{p}_{\ell|\delta}^{\delta'} = \frac{w_\ell^*}{\sum_{\ell'\in\delta} w_{\ell'}^*}\ , \qquad \forall\ \ell\in\delta\ ,\ \forall\ \delta\subseteq\delta',\ \forall\ \delta'\in\Delta$$

$$\Rightarrow \qquad \bar{p}_{\ell|\delta}^{\delta'} = p_\ell^{\delta} \qquad \forall\ \ell\in\delta\ ,\ \forall\ \delta\subseteq\delta',\ \forall\ \delta'\in\Delta\ .$$

On remarque donc que la probabilité de prendre une ligne donnée à l'intérieur d'un sous-ensemble donné est <u>indépendante</u> du type de l'usager (en autant que celui-ci contienne le sous-ensemble choisi). Ceci nous permettra, dans la suite, d'interpréter une probabilité de répartition p_ℓ^δ soit domme la probabilité qu'un usager de type δ utilise la ligne ℓ, soit comme la probabilité conditionnelle de prendre la ligne ℓ à l'intérieur du sous-ensemble δ (et ce pour un usager de n'importe quel type δ' contenant δ). Nous nous servirons de cette propriété dans le chapitre huitième pour comparer les répartitions calculées à partir de notre sous-modèle à celles obtenues par simulation.

7.5 Calcul des temps d'attente

7.5.1 Introduction

La détermination du temps d'attente moyen pour un type d'usagers donné à un arrêt multiple est une tâche beaucoup plus complexe que le calcul du temps d'attente à un arrêt simple. Dans le chapitre cinquième, nous avons vu que le temps d'attente à un arrêt simple dépendait notamment du paramètre de forme de la distribution d'Erlang choisie pour décrire les intervalles de service à cet arrêt ; il paraît donc naturel de s'interroger d'abord sur la nature des intervalles de service à un arrêt multiple, du point de vue d'un type d'usagers donné. Après avoir étudié la distribution du minimum de n variables distribuées selon des lois d'Erlang, nous serons en mesure de calculer les moments de la distribution de l'intervalle de temps séparant l'arrivée d'un usager d'un type donné du passage du premier autobus d'une des lignes attrayantes pour cet usager ; ensuite, nous déduirons de ces moments, une "pseudo-distribution" pour les intervalles de service pour ce type d'usagers (nous parlons ici de "pseudo-distribution", car, en réalité, les intervalles de service successifs pour un type donné ne sont pas indépendants et ils ne suivent pas une distribution commune). A partir de la "pseudo-distribution" définie pour les intervalles de service, nous construirons une quasi-file d'attente pour chaque type d'usagers et nous utiliserons le temps d'attente moyen dans cette quasi-file comme approximation du temps d'attente des usagers du type concerné, ce qui nous permettra d'utiliser les résultats obtenus dans le chapitre cinquième.

7.5.2 Distribution du minimum de n variables suivant des lois d'Erlang

Dans cette sous-section, nous allons déterminer les moments de la distribution du minimum de n variables aléatoires indépendantes distribuées selon des lois d'Erlang avec des paramètres différents. Ces résultats nous seront des plus utiles dans la sous-section suivante.

Considérons n variables aléatoires $X_1, X_2, \ldots, X_n$ indépendantes distribuées selon des lois d'Erlang, ayant respectivement pour paramètre de forme et pour moyenne k_j et $1/\mu_j$, $j=1,2,\ldots,n$. Si l'on pose $\eta_j = \frac{k_j}{\mu_j}$, $j=1,2,\ldots,n$, on peut alors écrire la fonction de densité et la fonction cumulative de la distribution de X_j sous la forme :

$$f_{X_j}(t) = \frac{\eta_j^{k_j}}{(k_j-1)!} t^{k_j-1} e^{-\eta_j t}, \qquad t \geq 0$$

$$F_{X_j}(t) = 1 - e^{-\eta_j t} \sum_{i=0}^{k_j-1} \frac{(\eta_j t)^i}{i!}, \qquad t \geq 0 .$$

Il s'agit maintenant de déterminer la distribution du minimum des n V.A., à savoir de

$$Y = \min \{X_1, \ldots, X_n\} .$$

On a pour tout $t \geq 0$

$$\Pr\{Y > t\} = \Pr\{X_1 > t, X_2 > t, \ldots, X_n > t\}$$

$$= \Pr\{X_1 > t\} \Pr\{X_2 > t\} \ldots \Pr\{X_n > t\}$$

(par l'indépendance des X_j)

$$\Pr\{Y > t\} = \prod_{j=1}^{n} (1 - F_{X_j}(t))$$

$$= \prod_{j=1}^{n} \left\{ e^{-\eta_j t} \sum_{i=0}^{k_j-1} \frac{(\eta_j t)^i}{i!} \right\}$$

$$\Pr\{Y > t\} = e^{-\left(\sum_{j=1}^{n} \eta_j\right)t} \prod_{j=1}^{n} \sum_{i=0}^{k_j-1} \frac{(\eta_j - t)^i}{i!}$$

En posant $\bar{\eta} = \sum_{j=1}^{n} \eta_j$, on obtient

$$\Pr\{Y > t\} = e^{-\bar{\eta}t} \prod_{j=1}^{n} \left\{ \sum_{i=0}^{k_j-1} \frac{(\eta_j t)^i}{i!} \right\}$$

$$\Rightarrow \quad F_Y(t) = 1 - e^{-\bar{\eta}t} \prod_{j=1}^{n} \left\{ \sum_{i=0}^{k_j-1} \frac{(\eta_j - t)^i}{i!} \right\} , \quad t \geq 0 \qquad (10)$$

On peut remarquer ici que

$$\prod_{j=1}^{n} \sum_{i=0}^{k_j-1} \frac{(\eta_j t)^i}{i!} = \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \prod_{j=1}^{n} \frac{(\eta_j t)^{i_j}}{i_j!}$$

et donc

$$F_Y(t) = 1 - e^{-\bar{\eta}t} \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \prod_{j=1}^{n} \frac{(\eta_j t)^{i_j}}{i_j!} , \quad t \geq 0 \qquad (11)$$

On peut maintenant calculer les moments de la distribution de Y. Ces moments sont donnés par

$$E[Y^r] = \int_0^{\infty} t^r \, dF_Y(t) \qquad , r=1,2,\ldots$$

$$E[Y^r] = r \int_0^{\infty} t^{r-1} [1-F_Y(t)] \, dt$$

En remplaçant $F_Y(t)$ par sa valeur, on obtient

$$E[Y^r] = r \int_0^{\infty} t^{r-1} \; e^{-\bar{\eta}t} \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \prod_{j=1}^{n} \frac{(\eta_j t)^{i_j}}{i_j!} \, dt$$

Si l'on interchange l'ordre de l'intégrale et des sommations, on obtient après avoir regroupé les termes semblables

$$E[Y^r] = r \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \left(\prod_{j=1}^{n} \frac{\eta_j^{i_j}}{i_j!} \right) \int_0^{\infty} e^{-\bar{\eta}t} \, t^{[(r-1)+\sum_{j=1}^{n} i_j]} \, dt$$

Posant $x = r + \sum_{j=1}^{n} i_j$

et $y = \bar{\eta}t$,

on obtient, après multiplication du numérateur et du dénominateur par $\bar{\eta}^x$,

$$\int_0^{\infty} e^{-\bar{\eta}t} t^{[(r-1)+\sum_{j=1}^{n} i_j]} \, dt = \frac{1}{\bar{\eta}^x} \int_0^{\infty} e^{-\bar{\eta}t} \, \bar{\eta}^{x-1} \, t^{x-1} \, \bar{\eta} \, dt$$

$$= \frac{1}{\bar{\eta}^x} \int_0^{\infty} e^{-y} \, y^{x-1} \, dy$$

$$= \frac{\Gamma(x)}{\bar{\eta}^{x}}$$

$$= \frac{(x-1)!}{\bar{\eta}^{x}}$$

$$= \int_0^{\infty} e^{-\bar{\eta}t} t^{[(r-1)+\sum_{j=1}^{n} i_j]}\, dt = \frac{[(r+\sum_{j=1}^{n} i_j)-1]!}{\bar{\eta}^{r}\,\bar{\eta}^{\Sigma i_j}}$$

Ceci nous permet de conclure que

$$E[Y^r] = \frac{r}{\bar{\eta}^{r}} \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \left(\prod_{j=1}^{n} \frac{\eta_j^{i_j}}{i_j!} \right) \left(\frac{[(r-1)\sum_{j=1}^{n} i_j]!}{\bar{\eta}^{\Sigma i_j}} \right),$$

ce qui peut être réécrit sous la forme

$$E[Y^r] = \frac{r}{\bar{\eta}^{r}} \sum_{i_1=0}^{k_1-1} \left(\frac{\eta_1}{\bar{\eta}}\right)^{i_1} \frac{1}{i_1!} \sum_{i_2=0}^{k_2-1} \left(\frac{\eta_2}{\bar{\eta}}\right)^{i_2} \frac{1}{i_2!} \cdots \sum_{i_{m-1}=0}^{k_{m-1}-1} \left(\frac{\eta_{m-1}}{\bar{\eta}}\right)^{i_{n-1}} \frac{1}{i_{n-1}!}$$

$$\cdot \sum_{i_n=0}^{k_n-1} \left(\frac{\eta_n}{\bar{\eta}}\right)^{i_n} \frac{[(r-1)+\sum_{j=1}^{n} i_j]!}{i_n!}$$

ou, après avoir posé $\theta_j = \frac{\eta_j}{\bar{\eta}}$, $j=1,2,\ldots,n$, sous la forme

$$E[Y^r] = \frac{r}{\bar{\eta}^{r}} \sum_{i_1=0}^{k_1-1} \frac{\theta_1^{i_1}}{i_1!} \sum_{i_2=0}^{k_2-1} \frac{\theta_2^{i_2}}{i_2!} \cdots \sum_{i_{n-1}=0}^{k_{n-1}-1} \frac{\theta_{n-1}^{i_{n-1}}}{i_{n-1}!} \sum_{i_n=0}^{k_n-1} \frac{\theta_n^{i_n}}{i_n!} [(r-1)+\sum_{j=1}^{n} i_j]! \quad (12)$$

Il est clair que l'évaluation directe des moments de la distribution de Y par la relation (12) implique énormément de calculs, même pour des valeurs assez petites de n et des k_j. Il faut donc, si l'on désire évaluer ces moments, trouver une manière efficace d'organiser les calculs nécessaires.

Remarquons tout d'abord que

$$[(r-1) + \sum_{j=1}^{n} i_j]! = (r-1)! \prod_{\ell=1}^{n} \left\{ \frac{[(r-1) + \sum_{j=1}^{\ell} i_j]!}{[(r-1) + \sum_{j=1}^{\ell-1} i_j]!} \right\} \qquad (13)$$

et que

$$\frac{[(r-1) + \sum_{j=1}^{\ell} i_j]!}{[(r-1) + \sum_{j=1}^{\ell-1} i_j]!} = \prod_{m=1}^{i_\ell} [(r-1) + \sum_{j=1}^{\ell-1} i_j + m] \quad , \qquad \ell=1,\ldots,n \qquad (14)$$

(Note : pour $i_\ell = 0$, le dernier rapport est, bien sûr, égal à 1.)

En combinant (13) et (14), on obtient

$$[(r-1) + \sum_{j=1}^{n} i_j]! = (r-1)! \prod_{\ell=1}^{n} \left\{ \prod_{m=1}^{i_\ell} [(r-1) + \sum_{j=1}^{\ell-1} i_j + m] \right\} \qquad (15)$$

Ceci nous permet de réécrire (12) sous la forme

$$E\,[Y^r] = \frac{r!}{\bar{n}^r} \sum_{i_1=0}^{k_1-1} \frac{\theta_1^{i_1}}{i_1!} \left\{ \prod_{m=1}^{i_1} [(r-1)+m] \right\} \cdots \sum_{i_\ell=0}^{k_\ell-1} \frac{\theta^{i_\ell}}{i_\ell!} \left\{ \prod_{m=1}^{i_\ell} [(r-1)+\sum_{j=1}^{\ell-1} i_j + m] \right\}$$

$$\cdots \sum_{i_n=0}^{k_n-1} \frac{\theta_1^{i_1}}{i_1!} \left\{ \prod_{m=1}^{i_n} [(r-1) + \sum_{j=1}^{n-1} i_j + m] \right\} \qquad (16)$$

On peut remarquer aussi que pour ℓ quelconque dans $\{1,\ldots,n\}$,

$$\frac{\theta^{i_\ell}}{i_\ell!} = \prod_{m=1}^{i_\ell} [\frac{\theta}{m}] \qquad \text{si } i_\ell > 0 ,$$

et $$\frac{\theta^0}{0!} = 1 .$$

La relation (16) peut donc être mise sous la forme

$$E\,[Y^r] = \frac{r!}{\bar{n}^r} \sum_{i_1=0}^{k_1-1} \left\{ \prod_{m=1}^{i_1} \frac{\theta_1[(r-1)+m]}{m} \right\} \cdots \sum_{i_\ell=0}^{k_\ell-1} \left\{ \prod_{m=1}^{i_\ell} \frac{\theta_\ell[(r-1)+\sum_{j=1}^{\ell-1} i_j + m]}{m} \right\}$$

$$\cdots \sum_{i_n=0}^{k_n-1} \left\{ \prod_{m=1}^{i_n} \frac{\theta_n[(r-1)+\sum_{j=1}^{n-1} i_j + m]}{m} \right\} \qquad (17)$$

Si l'on pose, pour $\ell=1,\ldots,n$, $\sigma_\ell(\underset{\sim}{i}) = \sum_{j=1}^{\ell-1} i_j + (r-1)$, (17) devient

$$E\,[Y^r] = \frac{r!}{\bar{n}^r} \sum_{i_1=0}^{k_1-1} \left\{ \prod_{m=1}^{i_1} \frac{\theta_1[\sigma_1(\underset{\sim}{i})+m]}{m} \right\} \cdots \sum_{i_\ell=0}^{k_\ell-1} \left\{ \prod_{m=1}^{i_\ell} \frac{\theta_\ell[\sigma_\ell(\underset{\sim}{i})+m]}{m} \right\}$$

$$\cdots \sum_{i_n=0}^{k_n-1} \left\{ \prod_{m=1}^{i_n} \frac{\theta_n[\sigma_n(\underset{\sim}{i})+1]}{m} \right\} \qquad (18)$$

La relation (18) fait ressortir le fait que, pour évaluer une sommation donnée pour des valeurs données des indices des sommations "précédentes", il suffit de connaître la somme des valeurs de ces indices (et aussi, bien sûr, r).

Nous voudrions maintenant montrer que, si l'on fixe les valeurs des indices $i_1,\ldots,i_{\ell-1}$, on peut établir certaines relations entre les termes de la $\ell^{\text{ème}}$ sommation qui permettront de l'évaluer efficacement.

Le cas le plus simple est celui de la $n^{\text{ème}}$ sommation : $\sum_{i_n=0}^{k_n-1} \left\{ \prod_{m=1}^{i_n} \frac{\theta_n[\sigma_n(\underset{\sim}{i})+m]}{m} \right\}$. Si l'on appelle $\kappa_n(0),\ldots,\kappa_n(k_n-1)$, les valeurs des termes de cette sommation, on remarque que, pour $p \geq 1$,

$$\kappa_n(p) = \prod_{m=1}^{p} \frac{\theta_n[\sigma_n(\underset{\sim}{i})+m]}{m}$$

$$= \frac{\theta_n[\sigma_n(\underset{\sim}{i})+p]}{p} \prod_{m=1}^{p-1} \frac{\theta_n[\sigma_n(\underset{\sim}{i})+m]}{m}$$

$$\kappa_n(p) = \frac{\theta_n[\sigma_n(\underset{\sim}{i})+p]}{p} \kappa_n(p-1) \qquad (19)$$

Cette dernière relation nous permet de conclure que l'on peut évaluer cette sommation par factorisation partielle, en utilisant un algorithme semblable à la règle de Horner pour l'évaluation des polynômes.

Considérons maintenant une sommation quelconque qui ne soit pas la $n^{\text{ème}}$. Comme auparavant, nous supposons que les valeurs des indices

$i_1,\ldots,i_{\ell-1}$ sont fixées et nous appelons $\kappa_\ell(0),\ldots,\kappa_\ell(k_\ell-1)$, les valeurs des termes de la $\ell^{ème}$ sommation. Soit $S_{\ell+1}(\sigma_{\ell+1}(\underset{\sim}{j}))$ la valeur de la $(\ell+1)^{ème}$ sommation en fonction de la somme de ℓ premiers indices, on a

$$\kappa_\ell(p) = \left\{ \prod_{m=1}^{p} \frac{\theta_\ell [\sigma_\ell(\underset{\sim}{j})+m]}{m} \right\} S_{\ell+1}(\sigma_{\ell+1}(\underset{\sim}{j})), \quad p=0,\ldots,k_\ell-1$$

$$\Rightarrow \quad \kappa_\ell(p) = \left\{ \prod_{m=1}^{p} \frac{\theta_\ell [\sigma_\ell(\underset{\sim}{j})+m]}{m} \right\} S_{\ell+1}(\sigma_\ell(\underset{\sim}{j})+p)$$

$$\Rightarrow \quad \kappa_\ell(p) = \left\{ \prod_{m=1}^{p-1} \frac{\theta_\ell [\sigma_\ell(\underset{\sim}{j})+m]}{m} \right\} \frac{(\sigma_\ell(\underset{\sim}{j})+p)\theta_\ell}{p} S_{\ell+1}(\sigma_\ell(\underset{\sim}{j})+p), \quad p \geq 1$$

$$\text{Or,} \quad \kappa_\ell(p-1) = \left\{ \prod_{m=1}^{p-1} \frac{\theta_\ell [\sigma_\ell(\underset{\sim}{j})+m]}{m} \right\} S_{\ell+1}(\sigma_\ell(\underset{\sim}{j})+(p-1)), \quad p \geq 1 .$$

Les deux dernières relations font apparaître clairement l'existence de facteurs communs entre les termes successifs d'une même sommation. De plus, on voit apparaître une relation de récursivité entre les termes de la $\ell^{ème}$ sommation et les valeurs que prend la $(\ell+1)^{ème}$ sommation quand l'indice i_ℓ varie tandis que les indices i_1 à $i_{\ell-1}$ restent fixés.

A partir de ce que nous venons de trouver, il est possible de définir un algorithme récursif pour calculer $E[Y^r]$. Les "niveaux" de cet algorithme correspondent aux sommations. A chaque niveau, on calcule la sommation correspondante en fonction de la somme des indices précédents.

Pour simplifier l'exposition de l'algorithme, nous l'écrivons sous la forme d'une procédure SIMULA (ALGOL-60). Nous supposons dans

cette procédure que les tableaux globaux

integer array K [1:N]

et real array THETA [1:N]

et les variables globales

real TOTETA

et integer N, R

ont été déclarés et initialisés correctement.

```
real procedure VALSOMME (L,SIGMALI); integer L,SIGMALI;
    begin real SP, THETAL;
          integer  P;

        THETAL:=   THETA [L];

       comment traitement des valeurs 1 à k_ℓ-1 de l'indice i_ℓ;

       if   K [L] > 1   then
          begin   if  L < N
                  then begin
                     for P:= K [L]-1 step -1 until  1  do
                        SP:=(THETAL*(SIGMALI+P)/P)*(SP+VALSOMME(L+1,SIGMALI+P)
                  end
                  else begin
                     comment cas particulier de la n^ème sommation;
                     for  P:= K [N]-1 step -1 until 1 do
                        SP:= (THETAL*(SIGMALI+P/P)*(SP+1)
                  end
```

```
        end;
    comment traitement de la valeur 0 de i_ℓ;
    if  L < N
        then  VALSOMME:=SP+VALSOMME(L+1,SIGMALI)
        else  VALSOMME:=SP+1
end procedure VALSOMME;
```

Et pour calculer $E[Y^r]$, il suffit de faire

```
MOMENTR:=FACTORIELLE(R)*VALSOMME(1,R-1)/TOTETA**R
```

Une étude rapide de la complexité de l'algorithme que nous venons de définir nous montre que la procédure récursive sera appelée :

- 1 fois au niveau 1,
- k_1 fois au niveau 2,
- $k_1 k_2$ fois au niveau 3,
- $\prod_{j=1}^{\ell-1}$ fois au niveau ℓ,
- $\prod_{j=1}^{n-1} k_j$ fois au niveau n,

ce qui représente au total un nombre assez impressionnant d'appels. Pour chaque appel au niveau ℓ, on exécutera k_ℓ énoncés d'affectation de la forme "SP:=" ou "VALSOMME:=", ce qui veut dire qu'au total, on devra exécuter $\sum_{\ell=1}^{n} \left(\prod_{j=1}^{\ell} k_j \right)$ de ces énoncés pour calculer $E[Y^r]$. Ce total est

semblable à celui que l'on obtiendrait si l'on calculait le nombre d'énoncés d'affectation qu'il faudrait exécuter si l'on écrivait une procédure récursive pour évaluer l'expression (12) directement. La grande différence entre la procédure que nous proposons et la procédure "directe" se situe au niveau des calculs impliqués par chaque exécution d'un énoncé d'affectation :

- au plus 7 opérations primaires (4 additions, 2 multiplications et 1 division) pour VALSOMME,
- au moins une élévation à une puissance (calcul de θ^{i_ℓ}) et un calcul de factorielle ($i_\ell!$) plus quelques opérations élémentaires (additions pour la sommation et le passage des nouveaux paramètres au niveau "suivant") pour la procédure directe.

Il est donc clair que la procédure VALSOMME sera beaucoup plus rapide que la procédure "directe".

Il est important de souligner que l'on peut implanter l'algorithme proposé même dans un langage qui ne permet pas d'écrire des procédures récursives. Il suffit d'ajouter pour cela trois tableaux de longueur n donnant les valeurs des indices, des sommes d'indices et des sommes partielles (SP) à chaque niveau, ainsi qu'une variable indiquant laquelle des sommations on est en train de traiter.

7.5.3 Superposition de n processus d'Erlang

Considérons la situation d'un usager qui se présente à un arrêt multiple et pour lequel n des lignes desservant l'arrêt sont attrayantes.

Si, pour chacune de ces lignes attrayantes, les intervalles entre les autobus sont des V.A. indépendantes distribuées selon une même loi d'Erlang, on peut dire que le processus stochastique décrivant les passages des autobus attrayants pour notre usager constitue une superposition de processus d'Erlang. Chacun des processus originaux (associés aux lignes) constitue un processus de renouvellement, mais, en général, le processus résultant de leur superposition ne constitue par un processus de renouvellement (voir Feller (1966), p. 370, à ce sujet) ; il est donc inutile d'essayer de déterminer la distribution des intervalles séparant le passage de deux autobus. On peut cependant déterminer la distribution de l'intervalle de temps séparant l'arrivée d'un usager, se présentant au hasard, du passage du prochain autobus attrayant. Pour ce faire, nous allons appliquer la méthode des étapes ("stages") à nos processus d'Erlang originaux.

La méthode des étapes consiste à considérer une distribution d'Erlang de paramètre de forme k comme la somme de k distributions exponentielles de moyenne k fois plus petite. Si on appelle k_j' et $1/\mu_j$ le paramètre de forme et la moyenne de la distribution des intervalles de service de la $j^{ème}$ ligne attrayante ($j=1,\ldots,n$), on peut définir, à chaque instant, l'état du $j^{ème}$ processus original comme étant le nombre d'étapes exponentielles de moyenne $\eta_j = \frac{k_j'}{\mu_j}$ restant à "accomplir" avant le passage du prochain autobus de la ligne j. Si l'on considère les n processus ensemble, on peut décrire l'état du processus global par des n-uplets de la forme

$$(k_1, k_2, \ldots, k_n) \quad \text{avec} \quad k_j \in \{1, \ldots, k_j'\},\ j=1, \ldots, n$$

A un instant donné t, l'état X(t) du processus global nous indique donc combien il reste d'étapes exponentielles à accomplir avant la prochaine transition de chacun des processus originaux.

Il est clair que, dans le processus global, les transitions sont de la forme

$$(k_1, k_2, \ldots, k_j, \ldots, k_n) \to (k_1, k_2, \ldots, k_j-1, \ldots, k_n)$$

pour un j quelconque (avec l'exception que si $k_j = 1$, on passe à k_j').

Avec cette structure d'états, le processus global constitue une chaîne de Markov finie ergodique en temps continu. On peut donc affirmer qu'elle va tendre à l'infini (quand $t \to \infty$) vers un état d'équilibre stochastique. C'est cet état qui nous intéresse.

Posons

$$p_{k_1,k_2,\ldots,k_n}(t) = \Pr\{X(t)=\{(k_1,k_2,\ldots,k_n)\} .$$

Nous pouvons maintenant écrire les équations de Chapman-Kolmogorov de la chaîne. Après avoir regroupé ensemble tous les termes o(h), on obtient

$$p_{k_1,k_2,\ldots,k_n}(t+h) = (1- \sum_{j=1}^{n} \eta_j\, h)\, p_{k_1,k_2,\ldots,k_n}(t)$$

$$+ \sum_{j=1}^{n} h\, \eta_j\, p_{k_1,k_2,\ldots,k_{j-1},k_j+1,\, k_{j+1},\ldots,k_n}(t)$$

$$+ o(h)$$

(Rem : nous utilisons la convention $k_j+1=1$ si $k_j=k'_j$)

$$\Rightarrow \frac{p_{k_1,k_2,\ldots,k_n}(t+h)-p_{k_1,k_2,\ldots,k_n}(t)}{h} = -\left(\sum_{j=1}^{n} \eta_j\right) p_{k_1,k_2,\ldots,k_n}(t)$$

$$+ \sum_{j=1}^{n} \eta_j \, p_{k_1,k_2,\ldots,k_j+1,\ldots,k_n}(t)$$

$$+ o(h)/h$$

Si on laisse tendre h vers 0, on obtient

$$\frac{dp_{k_1,k_2,\ldots,k_n}(t)}{dt} = -\left(\sum_{j=1}^{n} \eta_j\right) p_{k_1,k_2,\ldots,k_n}(t) + \sum_{j=1}^{n} \eta_j \, p_{k_1,k_2,\ldots,k_{j+1},\ldots,k_n}(t) \qquad (20)$$

Prenons maintenant la limite quand $t \to \infty$ des deux côtés de l'équation (20).

On obtient :

- à gauche, $\frac{dp_{k_1,k_2,\ldots,k_n}(t)}{dt} = 0$ (équilibre à l'infini) ;

- à droite, les $p_{k_1,k_2,\ldots,k_n}(t)$ sont remplacées par les probabilités d'équilibre.

En définitive,

$$\left(\sum_{j=1}^{n} \eta_j\right) p_{k_1,k_2,\ldots,k_n} = \sum_{j=1}^{n} \eta_j \, p_{k_1,k_2,\ldots,k_j+1,\ldots,k_n}$$

(avec $k_j+1=1$, si $k_j=k'_j$)

et ce pour tous les états $(k_1,k_2,\ldots,k_n)$.

Il est clair qu'une solution de ce système d'équations est

$$p_{k_1,k_2,\ldots,k_n} = a, \qquad \forall\, k_1,k_2,\ldots,k_n$$

où a est choisi de manière arbitraire.

Comme, d'autre part, $\sum_{(k_1,k_2,\ldots,k_n)} p_{k_1,k_2,\ldots,k_n} = 1$, on peut cependant conclure que

$$p_{k_1,k_2,\ldots,k_n} = \frac{1}{\text{card}\,\{(k_1,k_2,\ldots,k_n)\}} = \frac{1}{\prod_{j=1}^{n} k'_j} \tag{21}$$

Cette distribution est valable à n'importe quel instant choisi au hasard et, en particulier, au moment où arrive un usager (si l'arrivée de celui-ci est indépendante des passages d'autobus).

Appelons maintenant Z la V.A. décrivant l'intervalle de temps séparant l'arrivée d'un usager au hasard du passage du prochain autobus d'une ligne attrayante. On a

$$\Pr\{Z \le t\} = \sum_{(k_1,k_2,\ldots,k_n)} p_{k_1,k_2,\ldots,k_n} \cdot \Pr\{Z \le t \mid X=(k_1,k_2,\ldots,k_n)\}$$

où X est l'état du système au moment de l'arrivée du client. Par (21), on a alors que

$$\Pr\{Z \le t\} = \sum_{(k_1,k_2,\ldots,k_n)} \frac{1}{\prod_{j=1}^{n} k'_j} \Pr\{Z \le t \mid X=(k_1,k_2,\ldots,k_n)\}$$

Remarquons ici que $Pr\{Z<t|X=(k_1,k_2,\ldots,k_n)\}$ est la probabilité qu'il y ait, dans un intervalle de temps t, au moins une transition d'un des processus d'Erlang originaux étant donné qu'il reste k_1 "étapes" exponentielles à accomplir pour le premier, k_2 pour le second, etc. Autrement dit, Z est distribué comme le minimum de n V.A. suivant des lois d'Erlang de paramètre de forme k_j et de moyenne $\frac{\eta_j}{k_j}$. Ceci nous permet d'écrire, en se référant aux résultats de la sous-section 7.5.2,

$$Pr\{Z<t|X=(k_1,k_2,\ldots,k_n)\} = 1-e^{-\bar{\eta}t} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j-1} \frac{(\eta_j t)^{i_j}}{i_j!} \right] , \quad t \geq 0$$

où $\bar{\eta} = \sum_{j=1}^{n} \eta_j$.

On a donc

$$Pr\{Z<t\} = \sum_{(k_1,k_2,\ldots,k_n)} \frac{1}{\prod_{j=1}^{n} k_j!} \left\{ 1-e^{-\bar{\eta}t} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j-1} \frac{(\eta_j t)^{i_j}}{i_j!} \right] \right\}$$

$$Pr\{Z<t\} = 1 - \frac{e^{-\bar{\eta}t}}{\sum_{j=1}^{n} k_j!} \sum_{k_1=1}^{k_1'} \sum_{k_2=1}^{k_2'} \cdots \sum_{k_n=1}^{k_n'} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j-1} \frac{(\eta_j t)^{i_j}}{i_j!} \right] \qquad (22)$$

On sait que $$\prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j-1} \frac{(\eta_j t)^{i_j}}{i_j!} \right] = \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \cdots \sum_{i_n=0}^{k_n-1} \prod_{j=1}^{n} \frac{(\eta_j t)^{i_j}}{i_j!} \qquad (23)$$

Si l'on substitue (23) dans (22) et si l'on réarrange l'ordre des sommations, on obtient

$$\Pr\{Z<t\} = 1 - \frac{e^{-\bar{\eta}t}}{\prod_{j=1}^{n} k_j'} \sum_{k_1=1}^{k_1'} \sum_{i_1=0}^{k_1-1} \sum_{k_2=1}^{k_2'} \sum_{i_2=0}^{k_2-1} \cdots \sum_{k_n=1}^{k_n'} \sum_{i_n=0}^{k_n-1} \prod_{j=1}^{n} \frac{(\eta_j t)^{i_j}}{i_j!}$$

$$= 1 - \frac{e^{-\bar{\eta}t}}{\prod_{j=1}^{n} k_j'} \prod_{j=1}^{n} \left[\sum_{k_j=1}^{k_j'} \sum_{i_j=0}^{k_j-1} \frac{(\eta_j t)^{i_j}}{i_j!} \right]$$

$$= 1 - \frac{e^{-\bar{\eta}t}}{\prod_{j=1}^{n} k_j'} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j'-1} \sum_{k_j=i_j+1}^{k_j'} \frac{(\eta_j t)^{i_j}}{i_j!} \right]$$

$$\Pr\{Z<t\} = 1 - \frac{e^{-\bar{\eta}t}}{\prod_{j=1}^{n} k_j'} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j'-1} (k_j'-i_j) \frac{(\eta_j t)^{i_j}}{i_j!} \right]$$

Et on peut conclure

$$F_Z(t) \equiv \Pr\{Z<t\} = 1 - e^{-\bar{\eta}t} \prod_{j=1}^{n} \left[\sum_{i_j=0}^{k_j'} \frac{(k_j'-i_j)}{k_j'} \frac{(\eta_j t)^{i_j}}{i_j!} \right] \qquad (24)$$

En comparant (24) à l'équation (10) de la sous-section précédente, on remarque que la seule différence entre F_Z et la distribution cumulative du minimum des n distributions originales est l'apparition d'un terme $\frac{(k_j'-i_j)}{k_j'}$ dans chacune des sommations. Ceci nous permet de déduire facilement une expression pour les moments de Z à partir des résultats de la sous-section précédente. Après avoir posé $\varsigma_j = \frac{\eta_j}{\bar{\eta}}$ on tire de (12)

$$E[Z^r] = \frac{r}{\overline{n}^r} \sum_{i_1=0}^{k_1'-1} \frac{(k_1'-i_1)}{k_1'} \cdot \frac{\theta_1^{i_1}}{i_1!} \cdots \sum_{i_n=0}^{k_n'-1} \frac{(k_n'-i_n)}{k_n'} \frac{\theta_n^{i}}{i_n!} \left[(r-1)+\sum_{j=1}^{n} i_j\right] : \quad (25)$$

Pour évaluer (25), nous utiliserons la même approche que dans la sous-section précédente. Par un cheminement analogue à celui qui nous avait permis d'écrire la relation (18), on montre que

$$E[Z^r] = \frac{r!}{\overline{n}^r} \sum_{i_1=0}^{k_1'-1} \left\{ \frac{(k_1'-i_1)}{k_1'} \prod_{m=1}^{i_1} \frac{\theta_1[\sigma_1(\underset{\sim}{i})+m]}{m} \right\}$$

$$\cdots \sum_{i_n=0}^{k_n'-1} \left\{ \frac{(k_n'-i_n)}{k_n'} \prod_{m=1}^{i_n} \frac{\theta_n[\sigma_n(\underset{\sim}{i})+m]}{m} \right\} \qquad (26)$$

où $\sigma_\ell(\underset{\sim}{i}) = \sum_{j=1}^{\ell-1} i_j + (r-1)$.

En poursuivant notre cheminement, nous sommes amenés à définir la procédure récursive VALSOMMEZ, calquée sur VALSOMME, pour calculer les moments de la distribution de Z. Si l'on suppose que les tableaux globaux

integer array K [1:N]

et real array THETA [1:N]

et les variables globales

real TOTETA

et integer N,R

ont été déclarés et correctement initialisés, la procédure prend la forme

```
real procedure VALSOMMEZ(L,SIGMALI); integer L,SIGMALI;
    begin  real  SP,THETAL;
           integer KL,P;
    THETAL:= THETA [L];
    KL:= K [L];
    comment traitement des valeurs 1 à k_l-1 de l'indice i_l;
    if  KL > 1  then
       begin if  L < N
             then begin
                for  P:=KL-1 step -1 until 1 do
                   SP:= (THETAL*(SIGMALI+P)/P)
                        *(SP+VALSOMMEZ(L+1,SIGMALI+P)*(KL-P))
             end
             else begin
                  comment cas particulier de la n^ème sommation;
                  for  P:= KL-1 step -1 until 1 do
                    SP:= (THETAL*(SIGMALI+P)/P)*(SP+KL-P)
             end
       end;
    comment traitement de la valeur 0 de i_l il faut diviser SP par KL;
    if  L < N
       then  VALSOMMEZ:=SP/KL+VALSOMMEZ(L+1,SIGMALI)
       else  VALSOMMEZ:= SP/KL+1
end procedure VALSOMMEZ;
```

Pour calculer $E[Z^r]$, il suffit de faire

```
MOMENTRZ:=FACTORIELLE(R)*VALSOMMEZ(1,R-1)/TOTETA**R
```

Il faut remarquer dans "VALSOMMEZ" qu'au lieu de multiplier à chaque itération le terme qu'on ajoute à la sommation partielle (c'est-à-dire soit VALSOMMEZ(L+1,SIGMALI+P), soit 1) par (KL-P)/KL, on ne fait que le multiplier par (KL-P) et l'on divise à la fin SP par KL, ce qui donne le même résultat, mais est plus efficace.

7.5.4 Pseudo-distribution induite pour les intervalles de service

Bien que, comme nous l'avons déjà fait remarquer, les intervalles séparant les passages d'autobus "attrayants" pour un type d'usagers donné à un arrêt multiple ne possèdent pas une distribution commune, il est possible de définir une "pseudo-distribution" qui approxime ce phénomène.

Considérons un type d'usagers δ choisi arbitrairement. Si on appelle μ_ℓ, la fréquence de la ligne ℓ, $\ell \in \delta$, la fréquence combinée des autobus attrayants pour le type δ est alors $\mu_\delta = \sum_{\ell \in \delta} \mu_\ell$; il est clair que l'intervalle de temps moyen séparant deux autobus attrayants est alors égal à $1/\mu_\delta$. Il paraît donc naturel d'exiger que la moyenne de la distribution choisie pour approximer le phénomène qui nous intéresse soit égale à $1/\mu_\delta$.

Si on désire utiliser une distribution d'Erlang comme pseudo-distribution des intervalles de service, il va falloir, en plus de la moyenne, que l'on détermine le paramètre de forme de celle-ci. Pour ce faire, nous allons utiliser les résultats de la sous-section précédente. Si on appelle

Z_δ la V.A. décrivant l'intervalle séparant l'arrivée d'un usager de type δ du passage d'un autobus d'une des lignes appartenant à δ et σ_δ^2 la valeur que l'on voudrait donner à la variance de notre pseudo-distribution, on peut établir la relation suivante :

$$E[Z_\delta] = \frac{1}{2\mu_\delta} + \frac{\sigma_\delta^2 \mu_\delta}{2} \qquad (27)$$

En effet, si les intervalles de service suivaient une distribution commune de moyenne $\frac{1}{\mu_\delta}$ et de variance σ_δ^2, on pourrait calculer $E[Z_\delta]$ par les relations donnant la vie résiduelle, c'est-à-dire par (27). Ici, essayons de procéder à l'inverse et d'utiliser (27) pour déterminer σ_δ^2 en fonction de μ_δ et $E[Z_\delta]$. D'autre part, si on utilise une distribution d'Erlang de paramètre de forme k_δ pour notre pseudo-distribution, on devra avoir

$$\sigma_\delta^2 = \frac{1}{k_\delta \mu_\delta^2} \qquad (28)$$

En substituant (28) dans (27), on obtient

$$E[Z_\delta] = \frac{1}{2\mu_\delta} + \frac{1}{2k_\delta \mu_\delta}$$

$$\Rightarrow \qquad k_\delta = \frac{1}{2\mu_\delta E[Z_\delta] - 1} \qquad (29)$$

En général, $[2\mu_\delta E[Z_\delta] - 1]^{-1}$ ne prendra par une valeur entière, alors que k_δ doit être entier. Il faut donc modifier légèrement (29) comme suit :

$$k_\delta = A\left[\frac{1}{2\mu_\delta\ E\,[Z_\delta] - 1}\right] \qquad (30)$$

où A(x) désigne l'entier le plus "proche" de x.

L'équation (30) nous donnera toujours une valeur "acceptable" (c'est-à-dire ≥ 1) pour k_δ, car il est certain que, pour tout δ, on aura $E\,[Z_\delta] \leq \frac{1}{\mu_\delta}$ (le cas-limite $E\,[Z_\delta] = 1/\mu_\delta$ correspond à la situation dans laquelle toutes les distribution des intervalles de service sont des exponentielles), ce qui implique que $(2\mu_\delta\ E\,[Z_\delta] - 1) < 1$. Remarquons aussi que le dénominateur $(2\mu_\delta\ E\,[Z_\delta] - 1)$ sera toujours positif, puisque, comme nous l'avons vu dans la sous-section 3.8.2 du chapitre troisième, on a toujours $E\,[Z_\delta] \geq (2\mu_\delta)^{-1}$ et l'égalité $E\,[Z_\delta] = (2\mu_\delta)^{-1}$ ne peut se produire que si les intervalles interautobus sont fixes et égaux pour toutes les lignes et si les déphasages entre les lignes sont optimaux.

En définitive, nous avons trouvé une manière de déterminer pour chaque type d'usagers une distribution d'Erlang qui approxime le processus de service tel que perçu par ce groupe d'usagers.

Soulignons, avant de terminer, que, si on utilise cette approche pour le modèle global, on peut déterminer les paramètres des pseudo-distributions pour tous les arrêts et pour toutes les décisions possibles des usagers dans une phase d'initialisation précédant la recherche proprement dite d'une affectation d'équilibre.

7.5.5 Paramètres à utiliser dans le calcul des temps d'attente

Si l'on désire utiliser le modèle de quasi-file d'attente $Q\text{-}M/E_k^{[K]}/1$ (et les résultats du chapitre cinquième) pour calculer les temps d'attente aux arrêts multiples, il va falloir être en mesure de préciser, pour chaque type d'usagers, les valeurs des quatre paramètres qui influent sur le temps d'attente :

- moyenne des intervalles de service,
- paramètre de forme de la distribution des intervalles de service,
- taille des groupes,
- coefficient d'achalandage du système.

Nous venons de voir dans la sous-section précédente comment l'on peut déterminer les deux premiers paramètres, c'est-à-dire $1/\mu_\delta$ et k_δ. Il ne nous reste donc plus qu'à trouver des valeurs adéquates pour la taille des groupes (que nous appellerons K_δ pour un type δ donné) et le coefficient d'achalandage (ρ_δ).

Intéressons-nous d'abord à ρ_δ. On peut penser à diverses mesures possibles pour représenter l'achalandage que subit un groupe d'usagers : moyenne des coefficients d'achalandage des lignes utilisées pondérée par les probabilités de répartition, rapport entre les taux d'arrivée pour le type d'intérêt (λ_δ) et la capacité résiduelle des lignes attrayantes, une fois que l'on a réparti les autres usagers, etc. La mesure que nous avons retenue, au lieu d'être "centrée" sur le type δ, vise plutôt à mettre en lumière l'achalandage global des lignes appartenant au sous-ensemble δ. Si on appelle $\underline{w}^*$ le vecteur des capacités résiduelles correspondant à la répartition d'équilibre pour l'arrêt, on a :

. capacité totale des lignes du sous-ensemble δ : $\sum_{\ell \in \delta} \nu_\ell$;

. capacité résiduelle des lignes du sous-ensemble δ : $\sum_{\ell \in \delta} w^*_\ell$;

. achalandage des lignes du sous-ensemble δ : $\sum_{\ell \in \delta} \nu_\ell - \sum_{\ell \in \delta} w^*_\ell$.

Nous sommes donc amenés à poser pour le coefficient d'achalandage ρ_δ

$$\rho_\delta = \frac{\sum_{\ell \in \delta} \nu_\ell - \sum_{\ell \in \delta} w^*_\ell}{\sum_{\ell \in \delta} \nu_\ell} = 1 - \frac{\sum_{\ell \in \delta} w^*_\ell}{\sum_{\ell \in \delta} \nu_\ell} . \qquad (31)$$

Il est intéressant de remarquer que, si δ ne possède qu'un élément (la ligne ℓ) et si aucun autre usager n'utilise la ligne ℓ, alors

$$\rho_\delta = \frac{\nu_\ell - w^*_\ell}{\nu_\ell} = \frac{\lambda_\delta}{\nu_\ell} = \frac{\lambda_\delta}{\bar{K}_\ell \mu_\ell} ,$$

ce qui correspond à la formule habituelle pour les systèmes d'attente avec service en groupes.

Il est certain que les autres mesures possibles que nous avons évoquées pour ρ_δ mériteraient que l'on s'y arrête plus longuement ; dans le cadre de ce travail, il paraît cependant préférable de s'en tenir à la formule (31) en soulignant les deux principaux avantages de celle-ci :

. mesure globale de l'achalandage des lignes attrayantes pour les usagers du type concerné ;

. calcul extrêmement simple de ρ_δ à partir du vecteur des capacités résiduelles fourni par le sous-modèle de répartition.

Passons maintenant à la détermination de K_δ. Dans ce modèle d'arrêt multiple, il parait extrêmement difficile de déterminer a priori K_δ. On peut en effet penser que K_δ doit refléter non seulement la capacité des autobus de chaque ligne et leur fréquence, mais aussi l'impact de la présence des autres usagers. En fait, on peut utiliser simplement pour K_δ la moyenne des capacités des autobus des lignes du sous-ensemble δ pondérée par la fréquence des lignes, car nous avons déjà tenu compte de la présence des usagers des autres types dans la définition de ρ_δ. On pose donc

$$K_\delta = \frac{\sum_{\ell \in \delta} \overline{K}_\ell \mu_\ell}{\sum_{\ell \in \delta} \mu_\ell}, \qquad (32)$$

$$\Longleftrightarrow \qquad K_\delta = \frac{\sum_{\ell \in \delta} \nu_\ell}{\mu_\delta}$$

Il est important de remarquer que les définitions de μ_δ, ρ_δ et K_δ procèdent d'une vision globale commune dans laquelle le temps d'attente pour un type donné δ dépend de la fréquence totale, de l'achalandage total et de la capacité totale des lignes du sous-ensemble δ.

Maintenant que nous nous sommes donnés des valeurs pour tous les paramètres de la quasi-file d'attente qui approxime le phénomène d'attente que vivent les usagers d'un type donné, il devient relativement simple

de calculer le temps d'attente moyen approximatif de des usagers en utilisant les résultats du chapitre cinquième. Il faut toutefois être attentif dans le choix des formules à employer. Dans le chapitre cinquième, certaines des formules découlent de l'emploi d'un taux d'arrivée normalisé à 1 (formule (70) par exemple) ; il ne faut pas utiliser ces formules en prenant comme valeur de λ λ_δ, car le modèle suggéré ici dépend du taux d'achalandage total des lignes du sous-ensemble δ (c'est-à-dire de $\sum_{\ell \in \delta} \nu_\ell - \sum_{\ell \in \delta} w_\ell$) plutôt que de λ_δ. Pour être certain de ne pas faire d'erreur, il est sans doute préférable d'utiliser la formule (71) du chapitre cinquième, car elle est définie en fonction des paramètres dont nous venons de préciser la valeur (μ_δ, k_δ et ρ_δ). On aurait alors pour le temps moyen d'attente approximatif pourles usagers de type δ :

$$\hat{W}_{q\delta}(K)s = \frac{1}{2\mu_\delta} \left\{ \left[1 + \frac{1}{k_\delta(1-\rho_\delta)} \right] + \frac{\rho_\delta}{K_\delta(1-\rho_\delta)} - \rho_\delta \left(1 - \frac{1}{K_\delta}\right) \left[\left(Z^*_{k_\delta} + \frac{1}{k_\delta}\right) + \left(Z^*_{k_\delta} - \frac{1}{k_\delta}\right)\rho_\delta^2 \right] \right\} \qquad (53)$$

7.6 Intégration du sous-modèle d'attente au modèle global d'affectation

Dans le modèle d'affectation des passagers pour les réseaux sans arrêts multiples, nous avions été en mesure d'intégrer sans difficulté les résultats fournis par le sous-modèle d'attente au modèle global par le biais des fonctions d'attente ; dans le cas du modèle pour les réseaux comportant des arrêts multiples, l'intégration sera plus ardue, car il faudra travailler tant au niveau des fonctions de coût sur les arcs d'attente qu'à celui des contraintes de répartition.

Il peut sembler difficile de réaliser, à première vue, comment on peut exprimer les probabilités de répartition en termes des variables du modèle global d'affectation, car le sous-modèle de répartition implique la résolution d'un système d'équations non-linéaires dans des variables qui ne sont pas définies dans le modèle global. En se penchant attentivement sur les définitions des variables dans le modèle global et le sous-modèle de répartition, on peut remarquer qu'il est possible de calculer la capacité résiduelle de chaque ligne "partant" d'un arrêt donné en faisant la différence entre la capacité totale de cette ligne et le flot total sur la section de cette ligne reliant l'arrêt en question à son successeur. On aurait donc pour un arrêt multiple r et une ligne ℓ ($\ell \in L_r^+$) donnés :

$$\begin{pmatrix} \text{capacité résiduelle} \\ \text{de la ligne } \ell \\ \text{après l'arrêt r} \end{pmatrix} = K_\ell - v^\ell_{r,\gamma^*_\ell(r)}$$

ou encore

$$w^*_\ell \quad (\text{pour l'arrêt r}) = K_\ell - v^\ell_{r,\gamma^*_\ell(r)} \tag{34}$$

Ceci dit, nous sommes maintenant en mesure de dresser une liste complète des correspondances existant entre les variables, paramètres et fonctions pour les deux modèles pour un arrêt multiple r donné.

Modèle d'arrêt multiple	Modèle global d'affectation
L^+	L_r^+
Δ	Δ_r
$\mu_\ell, \ell \in L^+$	$\mu_\ell, \ell \in L_r^+$
$k_\ell, \ell \in L^+$	$k_\ell(r), \ell \in L_r^+$
$\lambda_\delta, \delta \in \Delta$	$\tilde{v}_r^\delta, \delta \in \Delta_r$
$\bar{K}_\ell, \ \in L^+$	$(K_\ell - \bar{v}_r^\ell)/\mu_\ell, \ell \in L_r^+$
$\nu_\ell, \ell \in L^+$	$(K_\ell - \bar{v}_r^\ell), \quad \ell \in L_r^+$
$w_\ell^*, \ell \in L^+$	$(K_\ell - v_{r,\gamma_\ell^*(r)}^\ell), \ell \in L_r^+$
$\mu_\delta, \delta \in \Delta$	$\sum_{\ell \in \delta} \mu_\ell, \delta \in \Delta_r$
$p_\delta \ \delta \in \Delta$	$\dfrac{\sum_{\ell \in \delta} (v_{r,\gamma_\ell^*(r)}^\ell - \bar{v}_r^\ell)}{\sum_{\ell \in \delta} (K_\ell - \bar{v}_r^\ell)}, \delta \in \Delta_r$
$K_\delta, \delta \in \Delta$	$\dfrac{\sum_{\ell \in \delta} (K_\ell - \bar{v}_r^\ell)}{\sum_{\ell \in \delta} \mu_\ell}, \quad \delta \in \Delta_r$
$p_\ell^\delta, \ell \in \delta, \delta \in \Delta$	$p_{r,\ell}^\delta(\bar{\underline{v}}_r, \tilde{\underline{v}}_r), \ \ell \in \delta, \delta \in \Delta_r$
$\hat{W}_{q\delta}^{(K)s}, \delta \in \Delta$	$\Phi_r(\bar{\underline{v}}_r, \tilde{\underline{v}}_r), \quad \delta \in \Delta_r$

Tableau 7.1 - Correspondance entre les variables, paramètres et fonctions du modèle d'arrêt multiple et du modèle global d'affectation pour un arrêt multiple r donné.

La plupart des correspondances sont évidentes ; certaines demandent cependant quelques explications. Tout d'abord en ce qui concerne λ_δ et $\tilde{v}_r^\delta$, il suffit de remarquer que $\tilde{v}_r^\delta$ est le nombre moyen (par unité de temps) d'usagers prenant la décision δ à l'arrêt r, autrement dit, c'est le taux d'arrivée des usagers de type δ à l'arrêt, c'est-à-dire λ_δ. Quant à $\bar{K}_\ell$, c'est le nombre moyen de places libres (à l'arrêt r) dans les autobus de la ligne ℓ; dans le modèle de flot, K_ℓ est la capacité totale de la ligne ℓ et $\bar{v}_r^\ell$ est le nombre moyen de places occupées (par unité de temps), ce qui fait que $(K_\ell - \bar{v}_r^\ell)$ est le nombre moyen (par unité de temps) de places libres à l'arrêt r (c'est-à-dire ν_ℓ) et que $(K_\ell - \bar{v}_r^\ell)/\mu_\ell$ est le nombre moyen de places libres à l'arrêt r par autobus de la ligne ℓ (et on conclut que $\bar{K}_\ell = (K_\ell - \bar{v}_r^\ell)/\mu_\ell$)). Les expressions pour μ_ℓ et K_δ en termes des variables du modèle de flot sont obtenues en remplaçant dans les relations (31) et (32) de la sous-section 7.5.5 les w_ℓ^*, ν_ℓ, $\bar{K}_\ell$ et μ_ℓ par leurs équivalents dans le modèle de flot.

Il reste maintenant à exprimer, si possible, les probabilités de répartition en fonction des variables du modèle de flot. Dans la sous-section 7.4.7, on a montré que l'on avait

$$p_\ell^\delta = \frac{w_\ell^*}{\sum_{\ell' \in \delta} w_{\ell'}^*} \quad , \ell \in \delta, \ \delta \in \Delta \ .$$

En se servant de (34), on peut donc écrire

$$p^{\delta}_{r,\ell}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r) = \frac{K_\ell - v^{\ell}_{r,\gamma^*_\ell(r)}}{\sum\limits_{\ell'\in\delta}(K_{\ell'} - v^{\ell'}_{r,\gamma^*_{\ell'}(r)})} \quad , \quad \ell\in\delta,\ \delta\in\Delta_r$$

et l'équation (16) du système définissant les affectations réalisables (section 6.8) devient :

$$v^{+\delta d}_{r,\ell} \sum_{\ell'\in\delta}(K_{\ell'} - v^{\ell'}_{r,\gamma^+_{\ell'}(r)}) - \tilde{v}^{\delta d}_{r}(K_\ell - v^{\ell}_{r,\gamma^+_\ell(r)}) = 0 \ , \ d\in D, r\in R^m, \delta\in\Delta_r, \ell\in\delta \ .$$

Il est important de souligner que, si les lignes ne sont pas saturées, la dernière équation et les contraintes de conservation de flot aux sommets d'embarquement des arrêts multiples feront en sorte que les capacités résiduelles obtenues dans le modèle de flot satisferont les conditions que nous avons énoncées dans la sous-section 7.4.5 pour les répartitions d'équilibre.

En terminant, notons que pour calculer les $\Phi^{\delta}_{r}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)$, il suffit d'utiliser la relation (33) de la sous-section 7.5.5 en remplaçant μ_δ, ρ_δ et K_δ par leurs correspondants dans le modèle de flot et en prenant pour k_δ la valeur appropriée calculée dans la phase d'initialisation du modèle.

7.7 Quelques remarques sur l'approche employée

Dans le chapitre cinquième, nous avions été capables de définir un modèle théorique pour représenter le phénomène d'attente aux arrêts simples, ce qui nous permettait d'être certains de la "qualité" des résultats fournis par le modèle, si les hypothèses sur lesquelles il

reposait étaient vérifiés. Dans le cas des arrêts multiples, nous n'avons aucune certitude de ce genre puisque le modèle suggéré dans ce chapitre repose d'abord et avant tout sur un certain nombre d'"intuitions" sur la nature du phénomène à modéliser. Il est donc impossible d'évaluer a priori l'écart pouvant exister entre les résultats produits par le modèle proposé et les "véritables" valeurs des probabilités de répartition et des temps d'attente. Pour juger de la pertinence du modèle d'arrêt multiple dans sa forme actuelle, nous pourrons cependant effectuer des simulations d'arrêts et comparer les résultats du modèle à ceux obtenus par simulation. Dans le chapitre suivant, nous donnerons quelques comparaisons préliminaires à partir d'un petit nombre de simulations exploratoires.

CHAPITRE HUITIÈME

SIMULATIONS D'ARRÊTS

8.1 Introduction

Dans le chapitre précédent, nous avons suggéré un sous-modèle complexe pour déterminer de manière approximative la répartition et le temps d'attente moyen pour chaque type d'usagers à un arrêt multiple. De toute évidence, il est impossible de "qualifier" l'approximation produite par le sous-modèle proposé par des méthodes analytiques. La seule approche qui permette de se faire une idée sur la valeur du sous-modèle d'arrêt multiple consiste à simuler un arrêt et à comparer les répartitions et les temps d'attente "observés" dans les simulations aux valeurs prédites par le modèle. C'est ce que nous allons faire dans ce chapitre pour quelques exemples très simples d'arrêts multiples. Il doit être bien clair que les résultats présentés ici ne le sont qu'à titre indicatif et qu'ils ne constituent d'aucune façon une validation formelle du sous-modèle d'arrêt multiple. Une telle validation nécessiterait en effet un nombre considérable de simulations et l'emploi de techniques statistiques appropriées pour l'analyse des résultats (voir, par exemple, l'ouvrage de Bratley, Fox et Schrage (1983) à ce sujet). C'est donc dire que les résultats de ce chapitre ne peuvent servir qu'à suggérer des avenues de recherche possibles pour l'étude future du problème des arrêts multiples.

Dans le chapitre cinquième, nous avons décidé de négliger l'impact que pouvait avoir la variabilité de la taille des groupes de service sur

le temps d'attente moyen aux arrêts simples. Il nous a paru intéressant de réaliser quelques simulations d'arrêts simples pour déterminer l'importance de cet impact. Les résultats de ces simulations sont présentés dans la section 8.3 à titre indicatif comme les autres résultats de simulations.

8.2 Schéma général des simulations

L'algorithme général que nous avons utilisé pour les simulations est le suivant :

. au départ, le système est complètement vide ;

. on "réchauffe" le système pendant une certaine période, appelée "temps de réchauffement", sans prendre de mesures ;

. on continue la simulation pendant une période plus longue, appelée "temps de simulation", en notant, pour chaque usager qui prend l'autobus pendant cette période, la ligne à laquelle appartient l'autobus qu'il prend et le temps pendant lequel il a attendu ;

. on calcule pour chaque type d'usagers les probabilités de répartition simulées en divisant le nombre d'usagers de ce type qui ont utilisé chacune des lignes par le nombre d'usagers de ce type qui ont pris l'autobus pendant le "temps de simulation" ;

. on calcule pour chaque type d'usagers le temps moyen d'attente simulé en divisant le temps total d'attente des usagers de ce type (qui ont pris l'autobus pendant le "temps de simulation") par le nombre total d'usagers de ce type (qui ont pris l'autobus pendant le "temps de simulation").

Pour chaque cas qui nous intéressait, nous avons procédé à dix simulations distinctes de la situation et les résultats présentés correspondent aux moyenne de ces réplications (sauf dans les tableaux 8.3 et 8.4 où nous donnons les résultats de chaque réplication).

Dans toutes les simulations, les usagers sont engendrés selon des processus de Poisson indépendants (avec des germes aléatoires distincts).

A chaque fois que, dans un tableau, nous donnons les résultats comparatifs de plusieurs séries de simulations (tableaux 8.1, 8.3, 8.4, 8.5 et 8.6), nous avons utilisé les mêmes séries de nombres pseudo-aléatoires pour engendrer les arrivées des usagers et les intervalles inter-autobus.

Dans tous les exemples d'arrêts multiples, nous avons considéré des arrêts desservis par trois lignes et auxquels font accès des usagers dont le sous-ensemble de lignes attrayantes compte au moins deux lignes (il y a alors quatre types d'usagers : trois qui correspondent aux sous-ensembles de lignes de cardinalité 2 et un pour lequel toutes les lignes sont attrayantes).

8.3 Etude de l'impact de la variabilité de la taille des groupes

8.3.1 Aux arrêts simples

La formule d'approximation pour le temps d'attente moyen aux arrêts simples (chapitre cinquième) repose sur l'hypothèse que le temps d'attente moyen dans une quasi-file d'attente $Q\text{-}M/E_k^{[Y]}/1$ diffère très peu du temps d'attente moyen dans unq quasi-file $Q\text{-}M/E_k^{[K]}/1$ de mêmes paramètres

pour laquelle K = E [Y]. Il est certain qu'en général le nombre de places libres à bord des autobus d'une ligne desservant un arrêt donné varie ; il est donc important de voir si cette variabilité a un impact important sur le temps d'attente des usagers.

Pour mesurer l'importance de cet impact, nous avons décidé se simuler quatre arrêts en faisant varier la distribution du nombre de places libres à bord des autobus tout en gardant le nombre moyen de places libres par autobus fixé. Les quatre arrêts simulés correspondent à quatre distributions différentes pour les intervalles interautobus : exponentielle (Erlang-1), Erlang-2, Erlang-3 et déterministe (Erlang-∞). Pour chaque arrêt, nous avons considéré quatre cas : des autobus complètement vides, des autobus aux trois-quarts vides, des autobus à moitié vides et des autobus aux trois quarts pleins, en moyenne. La famille de distributions choisies pour le nombre de places libres par autobus est la binômiale, tel que suggéré par Kadosch (1976). Les résultats de ces simulations sont donnés dans le tableau 8.1.

Les résultats du tableau 8.1 sont assez faciles à interpréter : pour toutes les distributions des intervalles interautobus, le temps d'attente moyen des usagers est, à toutes fins pratiques, insensible à la variabilité de la taille des groupes, ce qui tendrait à confirmer l'hypothèse du chapitre cinquième.

Soulignons cependant que, dans toutes les simulations réalisées, la taille moyenne des groupes est telle que la probabilité d'avoir un groupe de taille nulle (autobus plein) est extrêmement faible. Il est

Distribution de Y	$\bar{W}_q^{sim}$ k=1	k=2	k=3	k=∞
déterministe ($\bar{Y}$=24)	38,293	29,465	25,194	10,270
binômiale [0,32] ($\bar{Y}$=24)	39,497	29,372	24,703	10,468
binômiale [0,48] ($\bar{Y}$=24)	38,695	31,844	27,800	11,285
binômiale [0,96] ($\bar{Y}$=24)	39,951	31,314	26,807	11,124
Moyenne	39,122	30,499	26,126	10,104
Écart maximum à la moyenne	0,829 (2,1%)	1,345 (4,4%)	1,674 (6,4%)	0,515 (4,8%)
Paramètre des simulations	Nombre de réplications pour chaque cas : 10 Temps de réchauffement : 100 Temps de simulation : 500 λ = 1 $1/\mu$ = 19,2 ρ = 0,8			

Tableau 8.1 : Impact de la variabilité de la taille des groupes sur le temps d'attente simulé moyen pour des quasi-files d'attente Q-M/$E_k^{[Y]}$/1

loin d'être certain que, dans le cas où il existe une probabilité assez forte d'avoir des groupes de taille nulle, l'impact de la variabilité de la taille des groupes soit négligeable ; en effet, dans un tel cas, on se trouve à modifier en quelque sorte la distribution des intervalles de service effectif, ce qui devrait influer sur le temps d'attente moyen. Il faut donc se garder de voir dans les résultats du tableau 8.1 une confirmation de notre hypothèse pour toutes les situations possibles de quasi-files d'attente avec service en groupes de taille variable.

8.3.2 Aux arrêts multiples

Après avoir vérifié le peu d'impact de la variabilité de la taille des groupes sur les temps d'attente aux arrêts simples, il convient de s'assurer que cette propriété tient aussi dans le cas des arrêts multiples. A cause des calculs exigés par les simulations d'arrêts multiples, nous nous sommes limités à la comparaison de deux séries de simulations. Dans la première série, les autobus ont des capacités fixes, tandis que, dans la seconde, les autobus sont, en moyenne, à moitié pleins. Comme pour les simulations d'arrêts simples, les distributions utilisées pour le nombre de places libres à bord des autobus sont des lois binômiales.

On trouvera dans le tableau 8.2 les paramètres des deux séries de simulations. Les tableaux 8.3 et 8.4 permettent de comparer les répartitions et les temps d'attente moyens simulés pour chacune des réplications des deux séries. Rappelons ici que les germes aléatoires utilisés pour des réplications de même nom sont les mêmes dans chaque série.

Paramètres communs aux deux séries	Nombre de réplications : 10 Temps de réchauffement : 20 Temps de simulation : 100 Distribution des intervalles de service : exponentielles Moyenne des intervalles de service : $1/\mu_1 = 4,0$ $1/\mu_2 = 1,5$ $1/\mu_3 = 1,0$ Taux d'arrivée : $\lambda_{12} = \lambda_{13} = \lambda_{23} = 20$ $\lambda_{123} = 40$
Paramètres spécifiques à la série 1	Capacités des autobus fixes Lignes 1 et 2 : 60 places par autobus Ligne 3 : 65 places par autobus
Paramètres spécifiques à la série 2	Capacités variant selon des lois binômiales Lignes 1 et 2 : binômiale [0,120], moyenne 60 Ligne 3 : binômiale [0,130], moyenne 65

Tableau 8.2 : Paramètres pour les séries 1 et 2 de simulations d'arrêt multiple

Probabilités de répartition	Série	Réplications										Moyenne
		A	B	C	D	E	F	G	H	I	J	
p_1^{12}	1	0,29	0,21	0,21	0,29	0,20	0,27	0,32	0,27	0,34	0,22	0,262
	2	0,29	0,21	0,20	0,29	0,20	0,27	0,33	0,26	0,34	0,22	0,263
p_2^{12}	1	0,71	0,79	0,79	0,71	0,80	0,73	0,68	0,73	0,66	0,78	0,738
	2	0,71	0,79	0,80	0,71	0,80	0,73	0,67	0,72	0,66	0,78	0,737
p_1^{13}	1	0,11	0,16	0,17	0,23	0,09	0,18	0,23	0,11	0,15	0,14	0,157
	2	0,11	0,16	0,17	0,23	0,10	0,18	0,24	0,11	0,14	0,15	0,159
p_3^{13}	1	0,89	0,84	0,83	0,77	0,91	0,82	0,77	0,89	0,85	0,86	0,843
	2	0,89	0,84	0,83	0,77	0,90	0,82	0,76	0,89	0,86	0,85	0,841
p_2^{23}	1	0,30	0,44	0,38	0,37	0,33	0,45	0,36	0,35	0,37	0,32	0,367
	2	0,31	0,44	0,37	0,37	0,33	0,46	0,37	0,36	0,37	0,33	0,371
p_3^{23}	1	0,70	0,56	0,62	0,63	0,67	0,55	0,64	0,65	0,63	0,68	0,633
	2	0,69	0,56	0,63	0,63	0,67	0,54	0,63	0,64	0,63	0,67	0,629
p_1^{123}	1	0,07	0,13	0,07	0,12	0,05	0,15	0,13	0,06	0,10	0,09	0,097
	2	0,07	0,12	0,08	0,11	0,05	0,15	0,13	0,06	0,10	0,08	0,095
p_2^{123}	1	0,29	0,36	0,33	0,31	0,34	0,38	0,30	0,32	0,33	0,27	0,323
	2	0,28	0,36	0,33	0,31	0,33	0,38	0,30	0,32	0,33	0,27	0,321
p_3^{123}	1	0,64	0,51	0,59	0,57	0,60	0,47	0,57	0,62	0,57	0,64	0,578
	2	0,65	0,51	0,59	0,58	0,61	0,47	0,56	0,62	0,57	0,65	0,581

Tableau 8.3 : Impact de la variabilité de la capacité sur les répartitions simulées pour un arrêt multiple donné

Temps d'attente moyens simulés	Série	Réplications A	B	C	D	E	F	G	H	I	J	Moyenne
$\bar{W}^{sim}_{12}$	1	3,78	1,77	2,59	1,83	4,34	1,54	1,48	3,63	1,56	4,42	2,694
	2	4,03	1,84	2,42	1,69	4,37	1,53	1,48	3,64	1,64	4,18	2,682
$\bar{W}^{sim}_{13}$	1	3,19	1,66	2,28	1,55	3,75	1,53	1,16	2,60	1,31	3,67	2,270
	2	3,41	1,70	2,16	1,38	3,77	1,52	1,15	2,62	1,35	3,39	2,245
$\bar{W}^{sim}_{23}$	1	2,85	1,42	1,93	1,32	3,46	1,30	0,94	2,40	1,06	3,66	2,034
	2	3,09	1,49	1,80	1,16	3,49	1,29	0,93	2,44	1,11	3,39	2,019
$\bar{W}^{sim}_{123}$	1	2,80	1,31	1,84	1,29	3,35	1,19	0,92	2,38	1,00	3,52	1,955
	2	3,05	1,37	1,71	1,11	3,39	1,18	0,89	2,36	1,04	3,26	1,936

Tableau 8.4 : Impact de la variabilité de la capacité sur les temps d'attente simulés moyens pour un arrêt multiple donné

A la vue des résultats des tableaux 8.3 et 8.4, il semblerait bien que l'impact de la variabilité de la capacité des autobus, tant sur les répartitions que sur les temps d'attente, soit tout à fait négligeable. Néanmoins, il n'est certainement pas possible d'en arriver à une conclusion ferme sur la base de ces seules simulations.

8.4 Questions relatives au sous-modèle de répartition

Le sous-modèle de répartition du chapitre septième a comme paramètre fondamental le taux de service des lignes desservant l'arrêt et on n'y tient par directement compte des fréquences des autobus sur les lignes. Cela nous amène à nous poser un certain nombre de questions.

La première de ces questions est la suivante : si l'on fait varier uniformément la capacité des autobus desservant un arrêt et la fréquence de ceux-ci de manière à garder constant le taux de service de chaque ligne, la répartition des usagers changera-t-elle? Pour répondre à cette question, nous avons construit à partir de la série 1 deux nouvelles séries de simulations : la série 3 dans laquelle les capacités des autobus sont cinq fois plus petites et les fréquences cinq fois plus grandes ; la série 4, dans laquelle on a des autobus deux fois plus gros, mais deux fois moins nombreux. Le tableau 8.5 permet de comparer les répartitions simulées moyennes pour ces trois séries. De toute évidence, les répartitions simulées sont très comparables pour les trois séries, ce qui nous amène à croire que la répartition des usagers est sans doute relativement insensible à des variations uniformes de la capacité des autobus (remarquons que

Série de simulations			1	3	4
Paramètres spécifiques à chaque série de simulations	Capacité des autobus de chaque ligne	1	60	12	120
		2	60	12	120
		3	65	13	130
	Fréquence des autobus sur chaque ligne	1	0,250	1,250	0,125
		2	0,667	3,333	0,333
		3	1,000	5,000	0,500
Moyenne des probabilités de répartition simulées	p_1^{12}		0,262	0,266	0,256
	p_2^{12}		0,738	0,734	0,744
	p_1^{13}		0,157	0,157	0,148
	p_2^{13}		0,843	0,843	0,852
	p_2^{23}		0,367	0,355	0,370
	p_3^{23}		0,633	0,645	0,630
	p_1^{123}		0,097	0,101	0,094
	p_2^{123}		0,323	0,309	0,328
	p_3^{123}		0,578	0,589	0,578

Paramètres communs aux trois séries	Nombre de réplications : 10 Temps de réchauffement : 20 Temps de simulation :100 Distribution des intervalles de service : exponentielles Taux d'arrivée : $\lambda_{12} = \lambda_{13} = \lambda_{23} = 20$; $\lambda_{123} = 40$

Tableau 8.5 : Impact d'une variation uniforme de la taille des autobus sur les probabilités de répartition.

l'on arriverait sûrement à une conclusion différente si l'on s'intéressait aux temps d'attente).

Notre seconde question peut se formuler comme suit : si l'on fait varier de manière différente les capacités et les fréquences des lignes tout en gardant les taux de service constants, la répartition des usagers sera-t-elle très différente?

Quant à la troisième question, elle est centrale à nos préoccupations : le sous-modèle de répartition au prorata des capacités résiduelles produit-il une "bonne" approximation de la répartition réelle des usagers?

Comme nous ne disposons par de données sur les répartitions réelles des usagers à des arrêts multiples, il nous faudra dans la dernière question remplacer les mots "répartition réelle" par "répartition simulée moyenne".

Pour répondre à nos deux dernières questions, nous avons décidé de construire, à partir de la série 1 de simulations, une nouvelle série dans laquelle, au lieu d'avoir des autobus de capacités comparables et des lignes avec des fréquences assez différentes, les fréquences de toutes les lignes sont égales et les capacités des autobus varient beaucoup d'une ligne à l'autre. Nous avons ensuite comparé les résultats du sous-modèle de répartition au prorata des capacités résiduelles aux répartitions simulées moyennes des deux séries de simulations et aux résultats que l'on aurait obtenus si on avait utilisé un modèle de répartition au prorata des fréquences (modèle qui a été utilisé dans la plupart des modèles antérieurs d'affectation des passagers comme nous l'avons vu dans le chapitre troisième). Tous ces résultats sont rassemblés dans le tableau 8.6.

Paramètres de service pour chaque cas	Résultats	Probabilités de répartition								
		p_1^{12}	p_2^{12}	p_1^{13}	p_{13}^{3}	p_2^{23}	p_3^{23}	p_1^{123}	p_2^{123}	p_3^{123}
Cas 1 Capacité des autobus ligne 1 : 60 ligne 2 : 60 ligne 3 : 65 Fréquences μ_1=0,250 ; μ_2=0,667 μ_3=1,000	obtenus par simulation (série 1)	0,262	0,738	0,157	0,843	0,367	0,633	0,097	0,323	0,578
	du modèle de répartition au prorata des fréquences	0,273	0,727	0,200	0,800	0,400	0,600	0,130	0,348	0,522
Cas 2 Capacité des autobus ligne 1 : 15 ligne 2 : 40 ligne 3 : 65 Fréquences $\mu_1 = \mu_2 = \mu_3 = 1$	obtenus par simulation (série 5)	0,319	0,681	0,241	0,759	0,388	0,612	0,081	0,345	0,572
	du modèle de répartition au prorata des fréquences	0,500	0,500	0,500	0,500	0,500	0,500	0,333	0,333	0,333
Cas 1 et 2	du modèle de répartition au prorata des capacités résiduelles	0,263	0,737	0,158	0,842	0,344	0,656	0,110	0,306	0,584
Taux d'arrivée des usagers $\lambda_{12} = \lambda_{13} = \lambda_{23} = 20$; $\lambda_{123} = 60$	Paramètres communs des simulations	Nombre de réplications : 10 Temps de réchauffement : 20 Temps de simulation : 100 Distribution des intervalles de service : exponentielle								

Tableau 8.6 : Comparaison des résultats obtenus par simulation avec ceux produits par les modèles de répartition au prorata des fréquences et des capacités résiduelles pour deux exemples d'arrêt multiple

Du tableau 8.6, il ressort plusieurs choses :

1) Une variation sensible des capacités et des fréquences peut provoquer des changements non-négligeables dans les répartitions simulées moyennes.

2) Quand les autobus sont de tailles comparables (série 1), la répartition produite par le sous-modèle proposé au chapitre septième est remarquablement proche de la répartition simulée moyenne.

3) Quand les autobus sont de tailles relativement différentes (série 5), on peut noter un certain écart entre la répartition produite par le sous-modèle proposé et la répartition simulée moyenne.

4) Dans les deux cas étudiés, les répartitions au prorata des capacités résiduelles sont nettement plus proches des répartitions simulées que les répartitions au prorata des fréquences.

Bien qu'on ne puisse conclure sur la base de deux exemples à la supériorité du sous-modèle de répartition au prorata des capacités résiduelles, les résultats du tableau 8.6 sont assez significatifs pour nous permettre d'affirmer que ce sous-modèle mérite une attention particulière et qu'il faudrait l'étudier plus à fond.

8.5 Résultats relatifs au sous-modèle d'attente

Par rapport au sous-modèle d'attente aux arrêts multiples, une seule question se pose : les temps d'attente prédits par le sous-modèle correspondent-ils aux temps d'attente simulés moyens?

Dans le tableau 8.7, nous comparons les temps d'attente de la série 1 de simulations à ceux obtenus en appliquant le sous-modèle d'attente proposé à la situation simulée. En analysant ces résultats, on peut remarquer que les temps d'attente prédits par le sous-modèle sont du même ordre de grandeur que les temps d'attente simulés, mais que les temps d'attente prédits ont tendance à être plus extrémistes que ceux obtenus par simulation. Il semblerait donc que le sous-modèle exacerbe les différences existant entre les types d'usagers.

Nous aurions pu établir des comparaisons semblables à celles du tableau 8.7 pour d'autres séries de simulations, mais nous avons préféré nous en abstenir parce que la variance des temps d'attente simulés est assez élevée, ce qui nous amène à penser qu'une étude sérieuse du sous-modèle d'attente nécessiterait un effort considérable sur le plan de l'analyse statistique des résultats des simulations.

8.6 Conclusion

Les résultats des simulations d'arrêts simples et d'arrêts multiples que nous avons effectuées sont extrêmement encourageants : de manière globale, il semble bien que les hypothèses à partir desquelles les sous-modèles des chapitres cinquième et septième ont été élaborés, soient raisonnables, et les répartitions et les temps d'attente prédits par les sous-modèles sont toujours relativement proches des valeurs obtenues dans les simulations.

Type d'usagers (δ)	{1,2}	{1,3}	{2,3}	{1,2,3}
Intervalle moyen interautobus $1/\mu_\delta$	1,09	0,80	0,60	0,52
Paramètre de forme de la pseudo-distribution des intervalles de service k_δ	1	1	1	1
Taille moyenne des autobus pour le sous-ensemble δ K_δ	60,00	64,00	63,00	62,61
Capacité totale des lignes du sous-ensemble δ	55,00	80,00	105,00	120,00
Capacité résiduelle des lignes du sous-ensemble δ	8,32	13,87	17,81	20,00
Coefficient d'achalandage du sous ensemble δ ρ_δ	0,849	0,827	0,830	0,833
Temps d'attente moyen prédit par le sous-modèle d'attente $\hat{w}_{q\delta}^{(K)s}$	3,793	2,437	1,867	1,652
Temps d'attente moyen simulé $\bar{w}_{q\delta}^{sim}$ (série 1)	2,694	2,270	2,034	1,955

Paramètres de l'exemple Distribution des intervalles de service : exponentielle

Capacité des autobus	Fréquence des lignes	Taux d'arrivée
ligne 1 : 60	$\mu_1 = 0,250$	$\lambda_{12}=\lambda_{13}=\lambda_{23}=20$
ligne 2 : 60	$\mu_2 = 0,667$	$\lambda_{123}=40$
ligne 3 : 65	$\mu_3 = 1,000$	

Pour les simulations : 10 réplications; temps de réchauffement : 20; temps de simulation : 100

Tableau 8.7 : Comparaison des résultats obtenus par simulation avec ceux produits par le sous-modèle d'attente pour un arrêt multiple.

Dans le futur, il serait très intéressant de réaliser une étude approfondie de l'approche que nous avons suggérée pour préciser ces résultats préliminaires.

CHAPITRE NEUVIÈME

PROPRIÉTÉS DU MODÈLE
GÉNÉRALISATIONS POSSIBLES

9.1 Introduction

Dans ce chapitre, nous nous interrogeons tout d'abord sur certaines propriétés fondamentales du modèle d'équilibre global défini dans le chapitre sixième : existence et unicité d'une solution d'équilibre pour le modèle global, possibilité de définir un algorithme efficace pour calculer une affectation d'équilibre. Par la suite, nous montrons que le modèle proposé peut être facilement modifié pour tenir compte de l'existence de plusieurs classes d'usagers (cf. discussion de la section 2.7) ou pour traiter le problème de l'affectation des passagers avec demandes variables (cf. section 2.6). Nous terminons le chapitre par une critique générale du modèle d'affectation proposé et par une liste de suggestions relatives à des recherches futures.

9.2 Existence et unicité d'une solution d'équilibre pour le modèle global

Il est clair que, pour qu'un modèle d'équilibre pour l'affectation des passagers présente un intérêt pratique, il est nécessaire de savoir quelles conditions doivent satisfaire les paramètres du modèle pour qu'il existe une affectation d'équilibre. De plus, on voudrait que cette affectation soit unique, car l'existence de plusieurs affectations d'équilibre soulève de nombreuses questions quant à l'interprétation qu'il faut faire des résultats produits par le modèle (entre autres, celle de la

détermination de la "bonne" affectation, c'est-à-dire de celle qui correspond le mieux au comportement des usagers).

Plusieurs auteurs (entre autres, Smith, 1979 ; Dafermos, 1980 ; Fang, 1980 ; Aashtiani et Magnanti, 1981) se sont penchés sur la question de l'existence et de l'unicité d'une solution d'équilibre pour le problème de l'affectation de trafic routier et le problème d'affectation multimodale sous certaines hypothèses. Il est cependant impossible de reprendre les résultats de ces auteurs pour les appliquer au modèle que nous avons défini et ce, pour deux raisons :

a) Les fonctions d'attente aux arrêts simples et aux arrêts multiples obtenues respectivement dans les chapitre cinquième et septième ne sont pas définies sur l'ensemble du domaine des affectations réalisables, au sens des chapitres quatrième et sixième. Plus précisément, il faut remarquer que, pour un arrêt simple r donné, $\Phi_r(\bar{v}_r,\tilde{v}_r)$ n'est définie que si le coefficient d'achalandage de l'arrêt est strictement inférieur à 1, c'est-à-dire si on a $\tilde{v}_r + \bar{v}_r < K_{\ell(r)}$ où $K_{\ell(r)}$ est la capacité de la ligne desservant l'arrêt r, et que, pour un arrêt multiple r donné, les fonctions d'attente $\Phi_r^{\delta}(\underline{\bar{v}}_r,\underline{\tilde{v}}_r)$ ne sont définies que si la condition de stationnarité du théorème 7.1 est satisfaite, c'est-à-dire si, pour tout sous-emsenble δ de de lignes desservant l'arrêt r, on a

$$\sum_{\delta' \subseteq \delta} \tilde{v}_r^{\delta'} < \sum_{\ell \in \delta} (K_\ell - \bar{v}_r^{\ell}).$$

b) Les contraintes de répartition aux arrêts multiples (équation (16) de la section 6.8) sont complètement différentes des contraintes que l'on retrouve dans les autres modèles d'affectation ; elles modifient donc radicalement la nature même du domaine des affectations réalisables.

Il semble probable que l'on puisse contourner sans trop de mal les difficultés que pose, pour une preuve d'existence et d'unicité d'une affectation d'équilibre, le domaine de définition restreint des fonctions d'attente. Pour ce qui est des problèmes engendrés par la présence des contraintes de répartition aux arrêts multiples, ils sont de nature beaucoup plus complexe et il n'est pas possible, dans le cadre de ce travail, de leur trouver une solution. C'est donc dire qu'il faudra, dans le futur, s'attaquer résolument à la tâche de spécifier la nature et les propriétés de l'ensemble des affectations réalisables pour le modèle global d'équilibre, tel que défini dans les chapitre sixième et septième, et que ce n'est qu'une fois cette tâche accomplie que l'on pourra étudier les questions de l'existence et de l'unicité d'une affectation d'équilibre.

9.3 Considérations algorithmiques

Il est sans doute prématuré, au moment où l'on ignore encore s'il existe une ou plusieurs affectations d'équilibre, de s'interroger sur la possibilité de définir un algorithme efficace pour calculer une affectation d'équilibre. Certaines remarques s'imposent cependant.

La première idée qui vienne à l'esprit dans la recherche d'un algorithme pour notre problème consiste tout simplement à essayer de modifier un algorithme pour le problème d'affectation du trafic avec structure de coûts asymétrique (Dafermos, 1982 ; Nguyen et Dupuis, 1981 ; Bertsekas et Gafni, 1982 ; Marcotte, 1983) afin de l'adapter à notre modèle. Malheureusement, certains caractères spécifiques du modèle, entre autres, les contraintes de répartition aux arrêts multiples, sont tels qu'il paraît impossible de croire qu'une adaptation de ce genre puisse se faire sans que l'on ait à surmonter de sérieux obstacles. Ceci nous amène à penser qu'il faudra sans doute développer une "machinerie" algorithmique originale pour résoudre notre problème. Un tel algorithme, dans lequel, bien sûr, on pourrait reprendre certaines des idées utilisées dans les algorithmes d'affectation du trafic et des notions tirées de la théorie des inéquations variationnelles (puisque, comme l'a souligné Dafermos (1980), plusieurs problèmes d'affectation peuvent être vus comme des inéquations variationnelles), devrait à notre avis être élaboré à partir d'une étude sérieuse de ce qui distingue le modèle d'affectation des passagers des autres modèles d'affectation avec structure de coûts asymétrique, à savoir principalement les contraintes de répartition aux arrêts multiples.

A l'heure qu'il est, il est évidemment beaucoup trop tôt pour présumer des résultats d'une telle étude ; il faut donc considérer les quelques remarques qui précèdent comme une simple indication d'une direction de recherche possiblement très prometteuse.

9.4 Généralisation du modèle aux classes multiples d'usagers

Dans la section 2.7 du chapitre deuxième, nous avions indiqué que la plupart des modèles d'affectation des passagers reposaient sur l'hypothèse d'homogénéité des usagers par rapport aux coûts, et, par la suite, quand nous avons défini notre modèle, nous avons aussi supposé que cette hypothèse était vérifiée. Dans le cas où les usagers diffèrent trop dans leur appréciation relative des désagréments découlant des diverses composantes de leur déplacement, on ne peut pas utiliser directement le modèle proposé dans les chapitres qui précèdent, car les résultats obtenus seraient faussés. Il faut donc modifier légèrement le modèle pour tenir compte de l'hétérogénéité des usagers. Pour ce faire, on peut procéder comme suit :

a) définir un certain nombre de classes d'usagers de sorte que, à l'intérieur de chaque classe, les usagers soient relativement homogènes ;

b) évaluer pour chaque classe obtenue un jeu de coefficients de coût α_a, α_b, α_m et possiblement α_c (voir sous-section 4.2.4) pour transformer les temps de parcours en coûts généralisés ;

c) séparer pour chaque paire origine-destination la demande totale de transport en commun entre les classes d'usagers ;

d) définir pour chaque arc du réseau éclaté une fonction de coût pour chaque classe d'usagers en utilisant les coefficients de coûts qui s'appliquent à cette classe (voir sous-section 4.3.8 et section 6.7) ;

e) établir une différenciation des usagers (voir section 2.9 et sous-section 4.3.7) qui tienne compte à la fois de la destination des usagers et de leur classe ;

f) réécrire les conditions définissant les affectations réalisables (sections 4.4 ou 6.8) en tenant compte du nouveau schéme de différenciation des usagers (ceci ne pose aucun problème, sauf que le nombre d'équations et d'inéquations du modèle sera multiplié par le nombre de classes d'usagers) ;

g) reformuler la condition d'équilibre du modèle (voir sous-section 4.5.2 et section 6.9).

Dans la reformulation de la condition d'équilibre, il faudra maintenant déterminer pour chaque destination et chaque classe d'usagers un sous-graphe des chemins les plus courts ou des stratégies optimales vers la destination en utilisant les fonctions de coût qui s'appliquent à cette classe. La condition d'équilibre consistera alors à exiger que les seuls arcs utilisés par les usagers de chaque classe vers chaque destination appartiennent au sous-graphe des chemins les plus courts ou des stratégies optimales vers cette destination pour cette classe d'usagers. Notons, en terminant, qu'il sera aussi possible de généraliser la notion de mesure de déséquilibre (sous-section 4.5.4) et les propriétés et théorèmes qui s'y rattachent.

9.5 Modèle avec demandes variables

Dans le modèle que nous avons élaboré au fil des chapitres précédents, nous avons toujours supposé que la demande de transport en commun était fixe et connue. Il serait maintenant intéressant de voir comment l'on pourrait généraliser le modèle dans le cas où, pour chaque paire origine-destination, la demande de transport est donnée par une fonction dont le paramètre est le vecteur des longueurs des chemins les plus courts ou le vecteur des espérances de coûts des stratégies optimales pour l'affectation en main. Dans un tel cas, il suffit de considérer que les g^{od} qui apparaissent dans le système définissant les affectations réalisables sont maintenant des variables et non des paramètres du modèle, et d'ajouter aux conditions que l'on avait déjà pour définir une affectation d'équilibre, une condition d'équilibre de la demande, condition qui peut s'énoncer sous la forme

$$g^{od} = G^{od}(\underline{\Pi}(\underline{v})), \qquad \forall\ d\in D,\ \forall\ o\in O,$$

où G^{od} est la fonction de demande pour la paire (o,d) et $\underline{\Pi}(\underline{v})$ est le vecteur des longueurs des chemins les plus courts ou des espérances de coûts des stratégies optimales pour l'affectation en main.

9.6 Critique générale du modèle

Au début de cette thèse, nous nous étions fixés comme but de développer un modèle d'affectation des passagers qui puisse traiter adéquatement des réseaux congestionnés. Au fil des pages, nous avons élaboré un modèle d'équilibre statique dans lequel on accorde une grande importance

à l'impact de la congestion sur les temps d'attente et sur les répartitions des usagers aux arrêts multiples. La plupart des hypothèses sur lesquelles repose le modèle (comportement égoïste des usagers, additivité des coûts, etc.) paraissent tout à fait raisonnables et elles sont utilisées dans la majorité des modèles d'affectation des passagers. Il existe néanmoins trois aspects principaux qui peuvent amener des questions qui méritent réponses.

a) Le choix d'élaborer un modèle statique, alors que le phénomène modélisé est de nature essentiellement dynamique. Comme nous l'avons déjà indiqué dans le chapitre deuxième, le recours à une modélisation statique est absolument nécessaire si l'on veut pouvoir traiter des réseaux de transport en commun réels ; d'ailleurs, on doit utiliser une modélisation statique pour plusieurs problèmes de nature similaire (affectation du trafic routier, affectation multi-modale, etc.).

b) L'adéquation du sous-modèle d'arrêt multiple au phénomène qu'il est sensé modéliser. Il est certain que le sous-modèle proposé dans le chapitre septième ne nous permet pas de prédire parfaitement les répartitions et les temps d'attente aux arrêts multiples, cependant les résultats du chapitre huitième, sans être définitifs, nous permettent d'espérer que ce soit le meilleur modèle d'arrêt multiple qui ait été défini jusqu'à maintenant.

c) La complexité du modèle global d'affectation quand le réseau comporte des arrêts multiples. Dans les sections 9.2 et 9.3, nous

avons vu que la présence des contraintes de répartition aux arrêts multiples rendait difficile l'analyse du modèle global d'affectation. Il nous semble toutefois que la complexité du modèle proposé ne tienne pas tellement à la modélisation choisie, mais plutôt qu'elle soit inhérente au problème d'affectation des passagers dans les réseaux congestionnés comportant des arrêts multiples.

En résumé, nous croyons que le modèle d'affectation des passagers développé dans cette thèse est très satisfaisant du point de vue de la modélisation du problème ; il restera à faire en sorte, dans le futur, qu'il puisse aussi être fructueux, c'est-à-dire utile dans la résolution de problèmes réels.

9.7 Quelques avenues de recherche pour le futur

Dans le début de ce chapitre et dans le chapitre précédent, nous avons déjà dégagé un certain nombre de directions de recherche pour le futur : étude de la question de l'existence et de l'unicité d'une affectation d'équilibre ; analyse de la structure mathématique du domaine des affectations réalisables ; mise au point d'un algorithme efficace pour le calcul d'une affectation d'équilibre ; validation formelle du sous-modèle d'arrêt multiple par simulation.

Du point de vue de la modélisation, on pourrait sans doute améliorer le modèle proposé si on réussissait à modifier le sous-modèle d'arrêt multiple de manière à tenir compte à la fois de la capacité résiduelle et de la fréquence des lignes dans la détermination des répartitions des

usagers. Nous croyons que cela ne devrait pas être trop difficile à accomplir.

Une dernière avenue de recherche intéressante : dans la sous-section 4.5.4, nous avions indiqué que le concept de mesure de déséquilibre était très proche de celui de fonction d'écart défini par Hearn (1982) ; il nous paraîtrait important d'examiner attentivement les différences qui existent entre les deux notions, tant d'un point de vue théorique (interprétations diverses de la notion d'équilibre) que d'un point de vue pratique (pour la formulation d'algorithmes).

A la lecture de la liste qui précède, il est clair que, malgré le travail déjà accompli, nous sommes bien loin d'avoir épuisé le sujet de l'affectation des passagers dans les réseaux de transport en commun.

CONCLUSION

Au début de cette thèse, nous nous étions donné comme objectif l'élaboration d'un modèle pour l'affectation des passagers pouvant traiter adéquatement les réseaux congestionnés comportant des arrêts multiples. Au fil des chapitres, nous avons réussi à définir un modèle global d'équilibre pour l'affectation des passagers, modèle dans lequel nous insistons sur l'impact qu'a la congestion du réseau de transport en commun sur les temps d'attente des usagers aux arrêts. Nous avons aussi défini un sous-modèle d'arrêt multiple qui, sans être parfait, permet de représenter de manière très satisfaisante les arrêts multiples dans le modèle global. Nous avons donc atteint l'objectif premier que nous nous étions fixé.

En plus du modèle d'affectation proprement dit, nous avons aussi obtenu des résultats fort intéressants relativement aux systèmes d'attente avec service en groupes, particulièrement en ce qui concerne les approximations du temps d'attente moyen dans les quasi-files d'attente avec service en groupes. Ces résultats sont d'autant plus importants du fait qu'ils peuvent être utilisés non seulement dans l'étude des réseaux de transport en commun, mais aussi dans une multitude d'autres contextes.

Il ne nous reste plus qu'à espérer que les recherches nécessaires à la mise en application du modèle que nous avons proposé se poursuivront et qu'elles seront fructueuses.

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma gratitude la plus vive à l'endroit de Monsieur le Professeur Michael Florian, mon directeur de recherche, pour l'attention qu'il m'a prodiguée tout au long de mes recherches, et pour la très grande patience dont il a fait preuve au cours de ces longues années. Je veux aussi remercier Monsieur le Professeur Robert Cléroux, qui a bien voulu devenir mon co-directeur de thèse, et Monsieur Nelson Maculan, professeur invité au Département d'informatique et de recherche opérationnelle de janvier 1982 à décembre 1983, qui a accepté de lire une bonne partie du manuscrit.

Je remercie également tous les étudiants du Département d'informatique et de recherche opérationnelle et du Centre de recherche sur les transports qui, par leurs commentaires judicieux ou leur support moral, m'ont aidé à faire de cette thèse ce qu'elle est.

Il me faut aussi remercier le Ministère de l'Education du Québec, le Centre de développement des transports — Transports Canada et l'Université de Montréal pour leur soutien financier.

Je désire également exprimer mes remerciements les plus sincères à mon épouse Johanne et à ma fille Catherine pour leur compréhension qui m'a permis de mener à bien mes travaux.

Je tiens finalement à souligner l'extrême attention qu'a apportée Madame Madeleine Achard à la dactylographie de ce document. Je l'en remercie vivement.

BIBLIOGRAPHIE

AASHTIANI, H.Z. et MAGNANTI, T.L. (1981), Equilibria on a Congested Transportation Network. Siam Journal on Algebraic and Discrete Methods 2, 213-226.

ABDULAAL, M. et LE BLANC, L.J. (1979), Methods for Combining Modal Split and Equilibrium Assignment Models. Transportation Science 13, 292-314.

ANDREASSON, I. (1976), A Method for the Analysis of Transit Networks, in 2nd European Congress on Operations Research (édité par M. Roubens). North-Holland, Amsterdam.

ANDREASSON, I. et HASSELSTROM, D. (1979), Assigning passengers to parallel routes in public transport. Document interne, Volvo Bus Corporation, Suède.

BAILEY, N.T.J. (1954), On Queuing Processes with Bulk Service. J. Royal Stat. Soc. (B) 16, 80-87.

BERTSEKAS, D.P. et GAFNI, E.M. (1982), Projection Methods for Variational Inequalities with Application to the Traffic Assignment Problem. Math. Prog. Study 17.

BRATLEY, P., FOX, B.L. et SCHRAGE, L.E. (1983), A Guide to Simulation. Springer-Verlag, New-York.

BURRELL, J. (1976), Multipath route assignment : A comparison of two methods, in Traffic Equilibrium Methods (édité par M. Florian). Lecture Notes in Economics and Mathematical Systems 118. Springer-Verlag, Berlin.

CHAPLEAU, R. (1974), Réseaux de transport en commun : structure informatique et affectation. Thèse de Ph.D., Département d'informatique, Université de Montréal.

CHRIQUI, C. (1974), Réseaux de transport en commun : les problèmes de cheminement et d'accès. Thèse de Ph.D., Département d'informatique, Université de Montréal.

CHRIQUI, C. et ROBILLARD, P. (1975), Common Bus Lines. Transportation Science 9, 115-121.

DAFERMOS, S. (1980), Traffic Equilibrium and Variational Inequalities. Transportation Science 14, No. 1.

DAFERMOS, S. (1982), Relaxation Algorithms for the General Asymmetric Traffic Equilibrium Problem. Transportation Science 16, No 2, 231-240.

DALY, A.J. (1973), TRANSEPT - A Multipath Public Transport Assignment Model, in PTRC Seminar Proceedings.

DIAL, R.B. (1967), Transit Pathfinder Algorithm. Highway Research Record 205, 67-85.

DIAL, R.B. (1971), A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration. Transportation Research 5, 83-111.

DOW, P. et VAN VLIET, D. (1979), Capacity Restrained Road Assignment. Traffic Engineering and Control 20, 296-305.

DOWNTON, F. (1955), Waiting Time in Bulk Service Queues. J. Royal Stat. Soc.(B) 17, 256-261.

DOWNTON, F. (1956), On Limiting Distributions Arising in Bulk Service Queues. J. Royal Stat. Soc. (B) 18, 265-274.

FANG, S. (1980), Fixed Point Models for Equilibrium Problems on Transportation Networks. Report No. 80-7, Department of Mathematics, University of Maryland.

FEARNSIDE, K. et DRAPER, D.P. (1971), Public Transport Assignment - A New Approach. Traffic Engineering and Control 13, 298-299.

FELLER, W. (1966), An Introduction to Probability Theory and its Applications. 2e édition, John Wiley, New-York.

FINCH, P.D. (1959), A Probability Limit Theorem with Application to a Generalization of Queuing Theory. Acta Math. Acad. Sci. Hungar. 10, 317-325.

FLORIAN, M. (1977), A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. Transportation Science 11, 166-179.

FLORIAN, M. (1981), Utility, Entropy and a "Paradox" of Traffic Flow. Transportation Research A 15A, 327-330.

FLORIAN, M.,CHAPLEAU, R., NGUYEN, S., ACHIM, C., JAMES-LEFEBVRE, L., GALARNEAU, S., LEFEBVRE, F. et FISK, C. (1979), Validation and Application of EMME : An Equilibrium Based Two-Mode Urban Transportation Planning Method. Transportation Research Record 728, 14-22.

FLORIAN, M. et FOX, B. (1976), On the Probabilistic Origin of Dial's Multipath Traffic Assignment Method. Transportation Research 10, 339-341.

FLORIAN, M. et NGUYEN, S. (1976a), Recent Experience with Equilibrium Methods for the Study of a Congested Urban Area, in Traffic Equilibrium Methods (édité par M. Florian). Lecture Notes in Economics and Mathematical Systems 118. Springer-Verlag, Berlin.

FLORIAN, M. et NGUYEN, S. (1976b), An Application and Validation of Equilibrium Trip Assignment Methods. Transportation Science 10, 374-389.

FORD, L.R. et FULKERSON, D.R. (1962), Flows in Networks. Princeton University Press, Princeton, N.J.

GENDREAU, M. (1977), Un modèle d'équilibre pour l'affectation des passagers dans les réseaux de transport en commun. Mémoire de M.Sc., Département d'informatique, Université de Montréal.

HASSELSTROM, D. (1981), Public Transportation Planning - A Mathematical Approach. Thèse de Ph.D., Department of Business Administration, University of Gothenburg, Suède.

HEARN, D.W. (1982), The Gap Function of a Convex Program. Operations Research Letters 1, 67-71.

HEARN, D.W., LAWPHONGPANICH, S. et NGUYEN, S. (1983), Convex Programming Formulations of the Asymmetric Traffic Assignment Problem. Research Report No. 83-12, Department of Industrial and Systems Engineering, University of Florida.

HILLIER, F.S. et LO, F.D. (1971), Table for Multiple-Server Queuing Systems Involving Erlang Distributions. Technical Report No. 31, Department of Operations Research, Stanford University.

HILLIER, F.S. et YU, O.S. (1981), Queuing Tables and Graphs. North-Holland, New-York.

HOLM, J. (1973), A Model for Allocation of Buses in a Route Network. The Institute of Mathematical Statistics and Operations Research, The Technical University of Denmark.

HOLROYD, E.M. et SCRAGGS, D.A. (1966), Waiting Times for Buses in Central London. Traffic Engineering and Control 8, 158-160.

KADOSCH, M. (1976), Temps d'attente dans les transports urbains en commun. R.A.I.R.O. 10, No 2, 37-54.

KENDALL, D.G. (1953), Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of Imbedded Markov Chains. Ann. Math. Stat. 24, 338-354.

KLEINROCK, L. (1975), Queuing Systems Vol. 1 : Theory. John Wiley, New-York.

KOTIAH, T.C.T., THOMPSON, J.W. et WAUGH, W.A. O'N. (1969), Use of Erlangian Distributions for Single-Server Queuing Systems. J. App. Prob. 6, 584-593.

LAMPKIN, W. et SAALMANS, P.D. (1967), The Design of Routes, Service Frequencies and Schedules for a Municipal Bus Undertaking : A Case Study. Operations Research Quarterly 18, 375-397.

LAST, A. et LEAK, S.E. (1976), Transept : A Bus Model. Traffic Engineering and Control 17, 14-17, 20.

LE CLERCQ, F. (1972), A Public Transport Assignment Method. Traffic Engineering and Control 14, 91-96.

LEYVRAZ, J.P. (1980), Le programme NOPTS et son modèle d'affectation : un aperçu critique. Institut des Transports et Planification, Ecole Polytechnique Fédérale de Lausanne.

MANDL, C.E. (1980), Evaluation and Optimization of Urban Public Transportation Networks. European Journal of Operational Research 5, 396-404.

MARCOTTE, P. (1983), A new algorithm for solving variational inequalities over polyhedra, with application to the traffic assignment problem. Publication #316, Centre de recherche sur les transports, Université de Montréal.

MOORE, E.F. (1957), The Shortest Path Through a Maze. International Symposium on the Theory of Switching Proceedings. Harvard University.

NGUYEN, S. et DUPUIS, C. (1981), Une méthode efficace de calcul d'un trafic d'équilibre dans le cas des coûts non-symétriques. Publication #205, Centre de recherche sur les transports, Université de Montréal.

PETERSEN, E.R. (1971), Bulk Queues with Random Batch Size : With Application to Railroad Modelling. Working Paper Series No. 71-3, Canadian Institute of Guided Ground Transport, Queen's University, Kingston, Ontario.

POWELL, W.B. (1981), Stochastic Delays in Transportation Terminals : New Results in the Theory and Application of Bulk Queues. Thèse de Ph.D., Department of Civil Engineering, M.I.T.

RAPP, M.H., MATTENBERGER, P., PIGUET, S. et ROBERT-GRANPIERRE, A. (1976), Interactive Graphic System for Transit Route Optimization. Transportation Research Record 619.

SCHEELE, S. (1976), A Mathematical Programming Model for the Combined Distribution-Assignment of Traffic by Bus and by Car. Department of Mathematical Statistics, University of Stockholm.

SHARP, G.P. (1973), Public Transit System Network Models : Consideration of Guideway Construction, Passenger Travel and Delay Time, and Vehicle Scheduling Costs. Thèse de Ph.D., School of Industrial and Systems Engineering, Georgia Institute of Technology.

SHEFFI, Y. et DAGANZO, C.F. (1978), Another "Paradox" of Traffic Flow. Transportation Research 12, 43-46.

SMITH, M.J. (1979), The Existence, Uniqueness and Stability of Traffic Equilibria. Transportation Research B 13B, 295-304.

SPIESS, H. (1983), On Optimal Route Choice Strategies in Transit Networks. Publication #286, Centre de recherche sur les transports, Université de Montréal.

WARDROP, J.G. (1952), Some Theoretical Aspects of Road Traffic Research. Proceedings of the Institute of Civil Engineers, Part II, 325-378.